

Univerzita Karlova

Filozofická fakulta

Ústav informačních studií a knihovnictví

Bakalářská práce

Radka Planková

Vizualizace jako metoda prezentace dat
Data visualization as a new way of presentation

Praha 2017

Vedoucí práce: Mgr. Dita Malečková

Poděkování:

Ráda bych touto cestou poděkovala Mgr. Ditě Malečkové za cenné rady, trpělivost a ochotu při vedení mé bakalářské práce. Dále bych chtěla poděkovat svým blízkým za podporu v průběhu studia.

Prohlášení:

Prohlašuji, že jsem tuto bakalářskou práci vypracovala samostatně, že jsem řádně citovala všechny použité prameny a literaturu a že práce nebyla využita v rámci jiného vysokoškolského studia či k získání jiného nebo stejného titulu.

V Praze, dne 28. července 2017

.....

Radka Planková

ABSTRAKT

Tato bakalářská práce pojednává o vizualizaci jako o metodě prezentace dat v grafické podobě. Cílem práce je potvrdit, že vizualizace dat je efektivní způsob, jakým můžeme prezentovat data, která jsou v surovém stavu nepřehledná. Je představena stručná historie vizualizace dat a uvedena problematika terminologie tohoto pojmu. Je popsán samotný vizualizační proces s důrazem na objasnění důvodů pro použití této metody prezentace dat a jsou představeny vybrané základní a pokročilé typy grafů. V oblasti prezentace vizualizovaných dat jsou zmapovány aktuální trendy. Práce dále obsahuje přehled několika užitečných nástrojů, které se používají pro tvorbu vizualizace různých typů dat. V praktické části je předveden proces tvorby vizualizace na základě veřejně dostupných dat pomocí vybraných softwarových nástrojů jako příklad toho, jak efektivní vizualizace funguje a k čemu slouží.

Klíčová slova:

Vizualizace, data, graf, storytelling, infografika, historie vizualizace, interaktivita

ABSTRACT

This bachelor thesis deals with a visualization as a method of graphic data presentation. The aim of the work is to confirm that data visualization is an effective way of presenting data, which may be confusing in its raw format. First, a short history of data visualization is introduced together with a basic terminology of this issue. That is followed by a description of the visualization process, focusing on clarifying the reasons leading to choosing this data presentation method. Furthermore, selected basic and advanced graphs are introduced. Current trends in the area of visualized data presentation are mapped. The thesis contains an overview of several useful tools used for creating the visualization of various types of data. The practical part demonstrates the visualization creation process based on publicly available data using selected software tools in order to demonstrate how effective visualization works and what are its uses.

Keywords:

Visualization, data, graph, storytelling, infographic, history of visualization, interaction

OBSAH

Předmluva.....	7
Úvod	9
1 Historie vizualizace dat.....	10
1.1 Nejstarší formy vizualizace a první použití grafů	10
1.2 Nové formy grafického znázornění	11
1.2.1 William Playfair.....	11
1.3 Počátek moderní grafiky	11
1.3.1 John Snow.....	12
1.4 Zlatý věk statistické grafiky	12
1.4.1 Charles Joseph Minard	13
1.4.2 Francis Galton.....	14
1.5 Období temna	14
1.6 Současnost	15
2 Definice a terminologie	16
2.1 Vizualizace dat.....	16
2.2 Vizualizace informací	18
2.3 Vizualní analýza	19
2.4 Přehled oblastí vizualizace	20
2.5 Cíle a účely vizualizace	21
3 Proces vizualizace	23
3.1 Efektivita vizualizace dat.....	25
3.2 Základní schémata vizualizace dat.....	26
3.3 Pokročilé grafy	28
4 Aktuální trendy v oblasti prezentace dat	29
4.1 Storytelling.....	29
4.1.1 Specifika storytellingu u vědeckých dat	32
4.2 Informační grafika	33
4.3 Vliv vizualizace na paměť.....	34
4.3.1 Vliv interakce na paměť.....	35
4.4 Přímá vizualizace	36
4.5 Síťové vizualizace	37
5 Nástroje k vizualizaci dat.....	39
5.1 Out-of-the-box software.....	40
5.2 Programování.....	40
6 Praktická část procesu vizualizace dat	42
6.1 Kde získáme data?	42
6.2 Ukázka vizualizace dat	43
6.3 Google Spreadsheets	45
6.4 Vizualizace v R.....	45
6.5 Vizualizace v Pythonu	47
6.6 Vizualizace pomocí Out-of-the-box software	50
6.6.1 Plotly.....	50
6.6.2 DataHero	51
6.6.3 Tableau Software.....	52
6.7 Shrnutí praktické části	53

7	Závěr	55
	Seznam použité literatury.....	56
	Seznam obrázků	61

Předmluva

Téma vizualizace dat je v dnešní době moderní informační společnosti vysoce aktuální. Lidstvo je zahlcené obrovským množstvím dat, které musí správně interpretovat, aby porozumělo světu, v němž žije. Vzhledem k této složité různorodosti dat je často nejlepším způsobem komunikovat data vizuálně.

Již minulost nám poskytla mnohé příběhy, ve kterých se ukázalo, jak správné použití grafů a jiných vizuálních prezentací dat dokáže být efektivní. Takto prezentované informace se dají pochopit rychleji, jsou srozumitelné a častěji také snáze zapamatovatelné než pouhý text či tabulka bez grafického zpracování. V současné době nelze přehlédnout časté použití infografiky jakožto marketingového nástroje a propojení vizualizace dat s uměním. Velká pozornost je také věnována fenoménu big data a dostupnosti otevřených dat. Dnes již nestačí pouze získat kvalitní data, důležitá je jejich správná prezentace publiku. Proces vizualizace spočívá ve spolupráci mnoha odborníků od vědeckých pracovníků přes grafické designéry, programátory až po umělce, marketéry či obchodníky. Vizuální prezentace dat či informací nás neustále obklopuje a její význam bude dále narůstat. Nástroje pro vizualizaci se vyvíjí stále rychleji, jsou dostupnější, uživatelsky přívětivější a snadnější na ovládání i bez nutnosti znalosti programování.

Mým záměrem bylo získat do této problematiky vhled a pokusit se najít způsob, jak data zpracovat a převést je do grafického zobrazení. Absolvovala jsem tedy kurz datové analytiky DataGirls a programování v R, který pořádá spolek Czechitas¹ a kurz Pyladies², kde jsem na naučila základy programování v Pythonu. Získané poznatky z kurzů jsem využila v praktické části své bakalářské práce.

K vypracování této bakalářské práce byly využity zdroje zejména zahraničních autorů v anglickém jazyce. V seznamu použité literatury jsou uvedeny tyto typy dokumentů: elektronické a tištěné monografie, články z časopisů, závěrečné vysokoškolské práce, příspěvky z konferencí.

¹ Czechitas z.s. pořádají pro ženy kurzy programování, tvorby webu, grafiky a datové analytiky. Jejich cílem je zvýšení digitální gramotnosti a získání více žen do IT.

² Pyladies pořádají pro ženy kurzy programování v jazyce Python a další aktivity v oblasti IT

Bibliografické reference jsou připraveny v souladu s 3. vydáním mezinárodní normy ISO 690:2010. Pro odkazování na bibliografické citace v textu je použit harvardský systém.

Práce má celkem 42 normostran, tj. 75 452 znaků.

Úvod

Tato bakalářská práce pojednává o vizualizaci jako o metodě prezentace dat v grafické podobě. Data, která jsou vyjádřena pomocí grafiky, můžeme jednodušeji vnímat, pochopit a hledat v nich souvislosti než ta, která jsou vyjádřena jen v číselné či textové podobě. V současné době produkuje lidstvo obrovské množství dat, které se každým rokem zvyšuje. Denně na naše smysly působí velké množství informací, které bychom měly umět efektivně zpracovávat a získat do nich vhled.

Vizualizace dat již dávno není výsadou vědeckých pracovníků, setkáváme se s ní v umění, médiích a dalších aspektech našeho života. Rozvoj v oblasti vizualizace dat spočívá nejen v čím dál větší míře dostupnosti dat, ale také vznikem nástrojů pro kreativnější práci s daty bez potřeby znalosti programování.

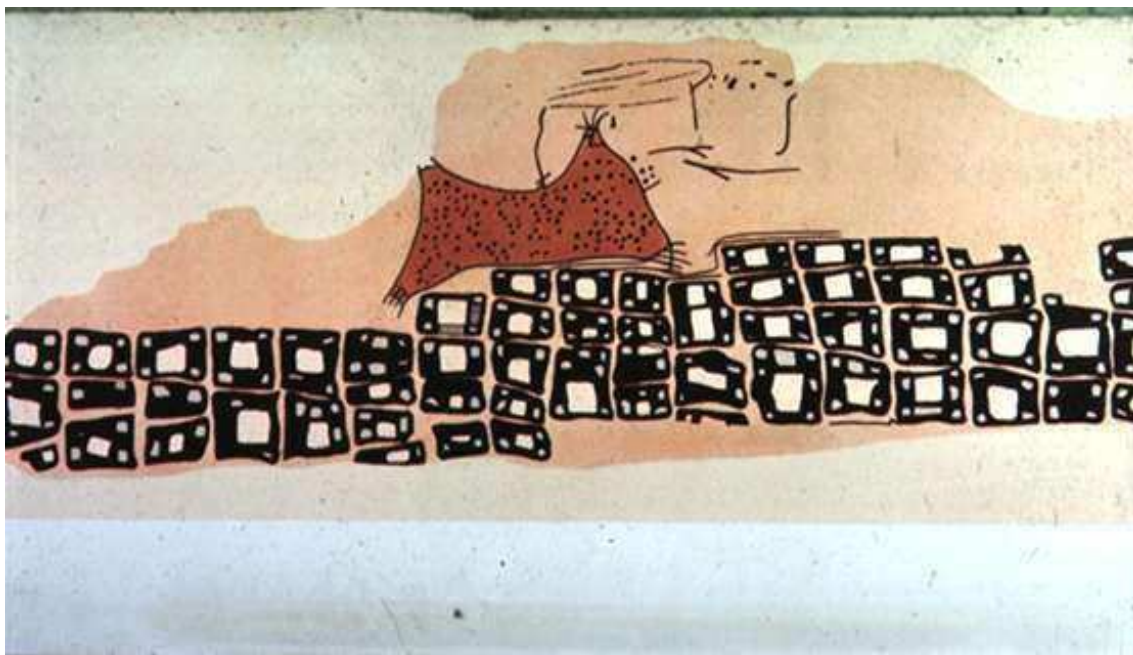
Bakalářská práce je rozdělena na teoretickou a praktickou část. Teoretická část obsahuje pět hlavních kapitol. V první kapitole je představena stručná historie vizualizace dat s důrazem na nejvýraznější milníky a významné osobnosti působící v oblasti grafického znázornění dat. Druhá kapitola se zabývá problematikou termínu vizualizace dat, který je nejednoznačný a obtížně definovatelný. Třetí kapitola se věnuje procesu vizualizace a jejímu efektivnímu fungování, jsou v ní také představeny základní a pokročilé typy grafů. Čtvrtá kapitola mapuje aktuální trendy a způsoby, jakými lze v dnešní době prezentovat a komunikovat data. Pátá kapitola poskytuje přehled několika užitečných nástrojů, pomocí kterých můžeme vytvářet vizualizace různých typů dat. V praktické části je předveden proces tvorby vizualizace na základě veřejně dostupných dat získaných z Českého statistického úřadu pomocí vybraných softwarových nástrojů. Cílem bakalářské práce je potvrdit, že vizualizace dat je efektivní způsob, jakým můžeme prezentovat data, která jsou v surovém stavu nepřehledná.

1 Historie vizualizace dat

Lidé obvykle považují statistickou grafiku a vizualizaci dat za poměrně novou metodu prezentace dat. „*Ve skutečnosti má však grafické znázornění kvantitativních informací hluboké historické kořeny.*“³ (Friendly, 2006, s. 1). Statistická grafika byla používána již v 17. století, jednoduché mapy a diagramy ještě v dřívějších dobách.

1.1 Nejstarší formy vizualizace a první použití grafů

Z historického hlediska není možné doložit první grafické znázornění dat. Nejstarší formy vizualizace nacházíme ve znázornění hvězdné oblohy a pozic nebeských těles a v mapách, které sloužily pro průzkum oblastí zemského povrchu a při zakládání nových měst. Nástěnná malba ve městě Konya v Turecku je považována za nejstarší mapu světa. Na této fresce je zobrazena část plánu města a dle letokruhů pochází z 8. st. př. Kr. (Saxl, Ilucová, 2004).



Obrázek 1: Rekonstrukce fresky nalezené v Catal Hüyük
zdroj: Saxl, Ilucová (2004)

Ve 14. století se v díle biskupa Mikuláše Oresme objevuje použití prvního sloupcového grafu, následně o století později vytváří významný renesanční myslitel Mikuláš Kusánský graf závislosti vzdálenosti na rychlosti. V 16. století jsou již techniky a nástroje pro pozorování a měření fyzikálních veličin, geografického zaměření a určení pozic

³ V původním znění: „*In fact, the graphic representation of quantitative information has deep roots.*“

nebeských těles dostatečně rozvinuty. Vznikají také první snímky vytvořené pomocí camera obscura nebo první moderní atlas. V 17. století je na vzestupu analytická geometrie, rozvíjí se teorie pravděpodobnosti, vědci se zabývají měřením času, vzdálenosti a prostoru. Jedná se také o období počátků demografické statistiky a systematického studia populace, půdy či zboží za účelem porozumění bohatství státu (Friendly, 2006).

1.2 Nové formy grafického znázornění

V 18. století dochází k rozmachu v oblasti nových forem grafického znázornění a reprezentaci dat. Pouhé znázornění geografické polohy na mapě již není dostačující. Objevují se proto první pokusy o tematické mapování ekonomických nebo lékařských údajů, vytváří se první abstraktní grafy a začínají se systematicky shromažďovat empirická data. (Friendly, 2006).

1.2.1 William Playfair

„Za zakladatele statistické grafiky je obecně považován William Playfaire“ (Saxl, Ilucová, 2004, s. 367). Je považován za vynálezce různých typů grafů široce používaných i v dnešní době – spojnicový, sloupcový a kruhový. (Friendly, 2006)

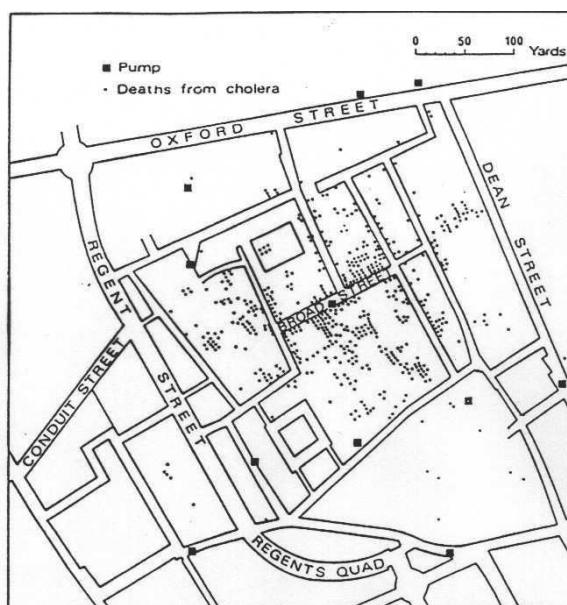
Playfaire se zajímal zejména o finance a obchod, proto se jeho nejvýznamnější práce zabývaly tématy národních financí, obchodu mezi různými státy, daňového zatížení a dalších. (Saxl, Ilucová, 2004).

1.3 Počátek moderní grafiky

Za počátek moderní grafiky považuje Friendly (2006) 19. století, kdy dochází, vzhledem k předchozím inovacím a zdokonalení techniky, k masivnímu nárůstu v počtu grafického zobrazení statistických dat a tematických map v takové míře, která nebyla až do moderních časů překonána. Ve statistické grafice byly v tomto období již všechny moderní formy zobrazení dat vynalezeny, také v oblasti tematické kartografie došlo k obrovskému rozvoji. Objevují se komplexní atlasy zobrazující údaje o nejrůznějších politických, ekonomických, demografických, lékařských a jiných tématech. Vědci ve svých publikacích začínají používat grafickou analýzu přírodních a fyzikálních jevů (Friendly, 2006).

1.3.1 John Snow

V tomto období působil také významný lékař John Snow, který byl průkopníkem v oblasti epidemiologických map. Snow pátral během epidemie cholery, která se objevila v Londýně v polovině 19. století, po zdroji infekce, a proto vytvořil mapu, která vizualizovala rozložení obětí této nemoci. Mapa se po svém vydání v roce 1854 stala velmi populární a Snow svým pečlivým pozorováním identifikoval zdroj kontaminace – byla to znečištěná voda ze studny na Broad Street. Jednalo se o skutečný průlom v medicíně, protože v té době panovalo přesvědčení o šíření cholery prostřednictvím vzduchu (Shiode, 2012).



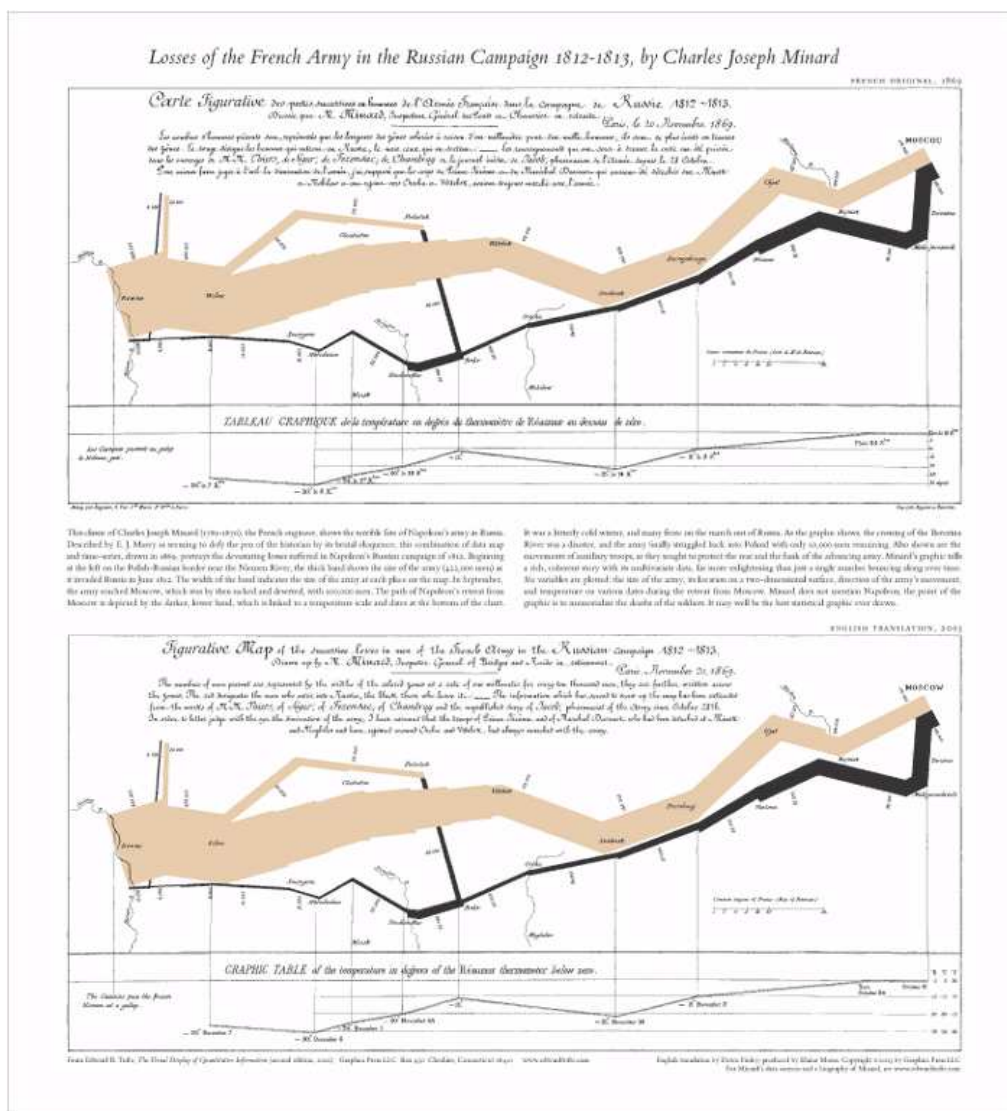
Obrázek 2: Snowův plán Londýna v době cholerové epidemie (1854)
zdroj: Saxl, Ilucová, (2004)

1.4 Zlatý věk statistické grafiky

Ve druhé polovině 19. století docházelo po celé Evropě ke zřizování oficiálních státních statistických úřadů. Důvodem byl rostoucí význam číselných informací v oblasti průmyslu, sociálního plánování nebo dopravy. Vývoj v oblastech tematické kartografie a grafiky je v této době velmi rozmanitý, zavádí se mnohé inovační prvky. Proto je toto období také nazýváno Zlatým věkem statistické grafiky (Friendly, 2006).

1.4.1 Charles Joseph Minard

Z roku 1869 pochází mapa od Charlese Josepha Minarda. Dle Tufteho (2013, s. 40) se jedná pravděpodobně „o nejlepší statistickou grafiku, která byla kdy vytvořena.“⁴ Minard vytvořil vývojový diagram o tragickém osudu Napoleonovy armády při tažení na Rusko v roce 1812.



Obrázek 3: Grafika Charlese Josepha Minarda, originál a anglický překlad
 zdroj: <https://www.edwardtufte.com/tufte/posters>

Zvláště účinným prvkem poskytujícím vysvětlení jevu je u zobrazování časových řad přidání prostorových rozměrů do konceptu grafiky tak, aby se data přesouvala jak prostorem, tak i časem. Při pohledu na tuto vizualizaci lze vidět, že šířka béžová linie značí v každém místě mapy velikost Napoleonovy armády. Cesta ústupu je značena

⁴ V původním znění: „It may well be the best statistical graphic ever drawn.“

černou linií, která je navázána na teplotní stupnici (všimněme si, že teploty byly výrazně pod bodem mrazu) s datem v dolní části obrázku.

Minardova grafika poskytuje bohatý a ucelený příběh, který sděluje a analyzuje vliv několika proměnných v čase (velikost armády, její lokace na dvourozměrném povrchu, směr vojenského tažení a teplota během jednotlivých dní). Grafika se tak stává mnohem více poučná než pouhé jediné číslo, které by se měnilo v čase (Tufté, 2013).

Přestože je Minard znám zejména jen pro tuto jednu práci, byl ve skutečnosti průkopníkem v oblasti tematické kartografie a statistické grafiky. Minard dával větší důraz na přesnou reprezentaci dat spíše než na přesnou geografickou polohu a představoval kvantitativní informace v numerické a vizuální podobě (Friendly, 2002).

1.4.2 Francis Galton

Další významnou osobností z konce 19. století je genetik a eugenik Francis Galton. Jeho zájem o grafickou reprezentaci statistických údajů vedl k rozvoji intenzivního zkoumání zejména v lékařství a genetice. Zavedl také pojmy regrese a korelační koeficient ((Saxl, Ilucová, 2004).

1.5 Období temna

Po období zlatého věku nastává začátkem 20. století temné období, kdy spíše než k rozvoji nových technik a metod ve zobrazování dat dochází k popularizaci a aplikaci statistické grafiky v mnoha vědních oborech. Píší se učebnice s podrobnými popisy grafických metod, které vedly k řadě nových poznatků a teorií. Jedná se o období očekávání nových technologií, které by podpořily další vlnu vývoje vizualizace dat (Friendly, 2006).

Probuzení z této letargie bylo způsobeno zejména vydáním průlomového dokumentu *The Future of Data Analysis* Johna W. Tukeyho (1962), pomocí něhož Tukey vyzýval k uznání analýzy dat jako legitimního odvětví statistiky odlišného od matematické statistiky. Tukey definoval explorační analýzu dat⁵, jakožto přístup k analýze dat, který využívá řadu různých technik a ukazuje, jak jednoduché grafické a kvantitativní techniky mohou být použity k průzkumu dat a hledání hypotéz k otestování. Explorační analýza

⁵ V původním znění: Exploratory Data Analysis (EDA)

dat slouží k nestrannému průzkumu dat před formulováním hypotéz, které se následně testují (konfirmační analýza). Tyto techniky umožní lepší posouzení předpokladů a výběr vhodných statistických nástrojů. Techniky EDA se také používají při dolování dat a k určení nejdůležitějších proměnných. (LIU, 2009).

1.6 Současnost

Největší rozvoj v oblasti vizualizace dat nastává od druhé poloviny 20. století vlivem rozvoje počítačové techniky a proniknutím grafiky do všech výzkumných oblastí. Pro každý počítač máme v současné době k dispozici širokou škálu softwarových nástrojů pro vizualizaci dat (Friendly, 2006).

Zvyšující se produkce grafů ale neznamená jejich rostoucí kvalitu či správnou interpretaci dat. Objevují se mnohé vizualizace, jejichž cílem jsou nejjednodušší formy reprezentace dat, pomocí kterých může autor snadno obrátit čtenářovu pozornost pouze k určité problematice bez komplexního pochopení problému (Saxl, Ilucová, 2004).

Důsledkem rychlého vývoje a používání vizualizačních metod v široké škále disciplín je velice obtížné poskytnout stručný přehled nejnovějšího vývoje v oblasti vizualizace dat. I přesto uvádí Friendly (2006) několik nejdůležitějších milníků: vývoj interaktivních statistických výpočetních systémů, nové metody pro vizualizaci velkorozměrových dat, kombinování různých metod analýzy dat pro překonání nedostatků jednotlivých technik, použití vizualizačních metod pro stále se rozšiřující řadu problémů a datových struktur, zvýšená pozornost ke vnímání zobrazovaných dat a další. (Friendly, 2006).

Vizualizací se v dnešní době zabývají nejen vědci v akademické sféře, ale její studium je dostupné všem také díky MOOC kurzům⁶. Velká pozornost je věnována fenoménu big data a hnutí za dostupnost dat pro širokou veřejnost (otevřená data). Některé z trendů v oblasti vizualizace jsou blíže popsány ve čtvrté kapitole.

⁶ z angl. výrazu *Massive Open Online Course*, v překladu znamená vzdělávací kurz pro neomezené množství účastníků probíhající online.

2 Definice a terminologie

Vizualizace dat nebo vizualizace informací jsou termíny, které se v současné době používají jak ve vědecké komunitě, tak postupně pronikají i do nevědeckých sfér. Termín vizualizace dat je nejednoznačný a obtížně definovatelný. Dle slovníkové definice znamená termín vizualizovat „*formovat mentální obraz něčeho, představovat si*“⁷. V literatuře autoři často používají termín vizualizace ne ve smyslu kognitivní aktivity dle definice výše, ale k popisu vizuální reprezentace dat.

2.1 Vizualizace dat

Card et al. (1999, s. 6) definuje vizualizaci dat jako „*používání počítačem podporovaných, interaktivních, vizuálních reprezentací dat za účelem zesílení poznávání*“⁸. Tato obecná definice zdůrazňuje vizualizaci jako práci s daty s cílem většího porozumění datům a tomu, co vlastně zobrazují.

Dle České terminologické databáze knihovnictví a informační vědy (dále jen TDKIV) je vizualizace informací, jejímž ekvivalentem je termín vizualizace dat „*proces převodu číselných a kvantitativních údajů a jejich vztahů do vizuálního, zpravidla grafického zobrazení s využitím počítačové grafiky, které napomáhá jejich percepci a porozumění. Obvyklými prostředky vizualizace jsou grafy, diagramy, sítě, mapy, grafické symboly, třírozměrné objekty, animace. Základními řešenými problémy jsou zobrazení vícerozměrných dat a výběr (filtrování) podstatných údajů pro vizualizaci.*“

Jako poslední si dovolíme uvést definici vizualizace jako „*proces zkoumání dat a informací po jejich převedení do grafické podoby. Jejím cílem je ... pochopení zkoumaných jevů a vniknutí do problému. Proto o vizualizaci mluvíme též jako o vizuální analýze dat*“ (Beneš, 1997, s. 6).

Vidíme, že výše uvedené definice mají i přes svou rozdílnost několik společných prvků. Zmiňují práci se surovými daty, pochopení vztahů mezi nimi a následné převedení do vizuální podoby. Tím se také liší vizualizace dat od prezentační grafiky. Ta pouze zprostředkovává informace, které jsou nám již dobře známé, na rozdíl od vizualizace,

⁷ V původním znění: „*Form a mental image of; imagine*“, slovníková definice viz.
<https://en.oxforddictionaries.com/definition/visualize>

⁸ V původním znění: „*the use of computer-supported, interactive, visual representations of data to amplify cognition*“

kteřá má za cíl nejprve pochopení daného jevu a vzájemných vztahů mezi nimi a následně zobrazení dat v grafické podobě. Nenapodobuje skutečnost, ale vytváří zcela nové obrazy jevů, které by pro nás v podobě surových dat zůstaly nepředstavitelné (Beneš, 1997).

Na chybějící obraz toho, co vizualizace představuje a na neexistující diskuzi ohledně přístupů k ní upozorňuje ve svém článku *Visualization Criticism – The Missing Link Between Information Visualization and Art* Robert Kosara. Uvádí, že zmatení kolem vizualizace pramení zejména z důvodu neexistence jasné a obecně přijatelné definice vizualizace. Definice by se dle něj měla mezioborově lišit, ale v rámci jedné oblasti by měla být konzistentní. Když porozumíme různým definicím vizualizace v rámci různých oborů, snadněji identifikujeme klíčové elementy, které mají jednotlivé obory společné (Kosara, 2007).

Kosara uvádí tři kritéria, která považuje za minimální množinu požadavků, abychom dílo mohli považovat za vizualizaci dat. Jako první kritérium uvádí, že vizualizace je založena na nevizuálních datech. Pokud tedy vstupním prvkem byla data ve formě obrazu, o vizualizaci se nejedná, protože výsledkem by byla produkce obrazu nikoli vizualizace dat. Druhým kritériem vizualizace je vytváření obrazu. Je zřejmé, že cílem každé vizualizace je převedení nevizuálních dat do vizuální podoby, a proto musí být vizuální zpracování, takové, které vnímáme zrakem, primárním prostředkem komunikace dat. Poslední kritérium se zabývá výsledkem procesu vizualizace. Výsledek musí být čitelný a rozpoznatelný, i když autor upozorňuje, že to vyžaduje trénink a praxi (Kosara, 2007).

Vizualizacemi vědeckých dat se zabývali již McCormick, DeFanti a Brown ve své práci *Visualization in Scientific Computing*, která vznikla na základě účasti na vědeckém panelu v roce 1987. Vizualizaci popisují jako počítačovou metodu, pomocí které převádíme data ze symbolického do geometrického tvaru. To pomáhá výzkumníkům sledovat jejich simulace a výpočty. Zdůrazňují také, že vizualizace nabízí metodu „*vidění neviděného*“⁹ a obohacují proces vědeckého poznávání, kdy podporují důkladný a mnohdy nečekaný vhléd do problematiky (McCormick et al., 1987).

Obvykle je vizualizace dat spojená s mylnou představou, že se jedná o shodný termín pro informační grafiku (infografika). Je důležité si uvědomit, že zatímco vizualizace dat vždy sděluje informaci pomocí objektivní vizualizace kvantifikovaných dat, infografika

⁹ V původním znění: „*seeing the unseen*“

se používá ke sdělování jakýchkoli informací – kvantifikovatelných i těch, které kvantifikovatelné nejsou (Laumans, 2009).

2.2 Vizualizace informací

Nejasně jsou také ukotveny rozdíly mezi pojmy vizualizace dat a vizualizace informací. V databázi TDKIV se jedná o ekvivalentní pojmy, stejně tak i někteří další autoři tyto pojmy nerozlišují. Card et al. (1999) u definice vizualizace informací přidává slovo abstraktní. Podle něj je tedy vizualizace informací „*používání počítačem podporovaných, interaktivních, vizuálních reprezentací **abstraktních** dat za účelem zesílení poznávání*“¹⁰ (Card et al., 1999, s. 7). Všimněme si, že obě definice sledují stejný účel, rozdílné jsou pouze vstupní prvky, tedy typy dat. U vizualizace (vědeckých) dat pracujeme s fyzicky ukotvenými daty, tudíž tvorba vizuální podoby je jednodušší a výsledek vizualizace bude pro publikum očekávanější.

Vizualizace informací je poměrně nová oblast výzkumu zaměřující se na takové využití vizualizačních technik, které pomáhají s porozuměním a analýzou dat. Zatímco vědecká vizualizace, jak je již výše uvedeno, pracuje s konkrétními prostorově ukotvenými daty, vizualizace informací se soustředí na abstraktní informace, které ve většině případů nelze mapovat do fyzického světa. Vlivem stále rostoucího množství generovaných dat, roste i poptávka po vizualizačních nástrojích. Vizualizace informací, stojící na pevných základech, v sobě kombinuje různé oblasti výzkumu od vědecké vizualizace, interakce člověk-počítač, dolování dat, informačního designu až po kognitivní psychologii či počítačovou grafiku (Kerren, 2008). Vizualizace informací tak osciluje mezi surovými daty a informacemi. Poskytuje metody a nástroje, pomocí kterých lze organizovat a reprezentovat data, aby získaly význam a produkovaly tak smysluplnou informaci (Mazza, 2009).

Podobně Telea (2015) rozlišuje mezi vědeckou vizualizací a vizualizací informací. Vědecká vizualizace dat se jako samostatná disciplína ustanovila v 80. letech minulého století, jako odpověď na stále větší množství počítačem generovaných simulací různých fyzických procesů. Slovo „vědecký“ zdůrazňuje zaměření se na takové řešení, které

¹⁰ V původním znění: „*the use of computer-supported, interactive, visual representations of abstract data to amplify cognition*“

poskytne vhléd do vědeckých simulací (Telea, 2015). V odborné literatuře se můžeme setkat také se zkráceným pojmem *scivis*.

Důležitým aspektem tohoto typu vizualizace je zobrazení souborů dat, které mají přirozené prostorové zakotvení, tedy elementy, které mají konkrétní význam v prostoru. Vzhledem k tomuto aspektu se v poslední době také setkáváme s termínem vizualizace prostorových dat¹¹ (Telea, 2015).

Kromě prostorových dat, existují i ostatní typy dat, které nemají předepsanou polohu v prostoru. Ačkoli tato data také potřebují určité prostorové zobrazení, abychom je mohli nakreslit či zobrazit na obrazovce počítače, prostorová informace je přiřazena k datovému prvku až v procesu vizualizace, nikoli v něm přímo obsažena (Telea, 2015). Tento typ vizualizace informací je zkráceně označován jako *infovis*.

2.3 Vizuální analýza

Tyto dva typy vizualizací doplňuje Telea (2015) novou oblastí – vizuální analýzou, která funguje jako rozšíření dvou původních oblastí vizualizace. Ta vznikla z potřeby kombinovat vizualizační řešení s datovou analýzou a dolováním dat. Ústředním cílem je poskytnutí technik a nástrojů, které podporují koncové uživatele v analytickém uvažování pomocí interaktivních vizuálních rozhraní.

Vizuální analýza je moderním přístupem kombinující lidskou intuici a matematickou dedukci sloužící k vnímání vzorů, odvození znalostí z nich a získání vhlédu. Vizuální analýza se zabývá tvorbou abstraktních vizuálních metafor kombinovanou interakcemi mezi lidmi a informacemi. Umožňuje ujištění očekávaného a objevení neočekávaného v rámci dynamicky se měnícího množství informací. Vizuální analýza vychází z oblasti vědecké a informační vizualizace, navíc zahrnuje technologie z mnoha dalších oblastí, včetně managementu znalostí, statistické analýzy, kognitivní vědy, vědy o rozhodování a mnoha dalších (Wong, 2004).

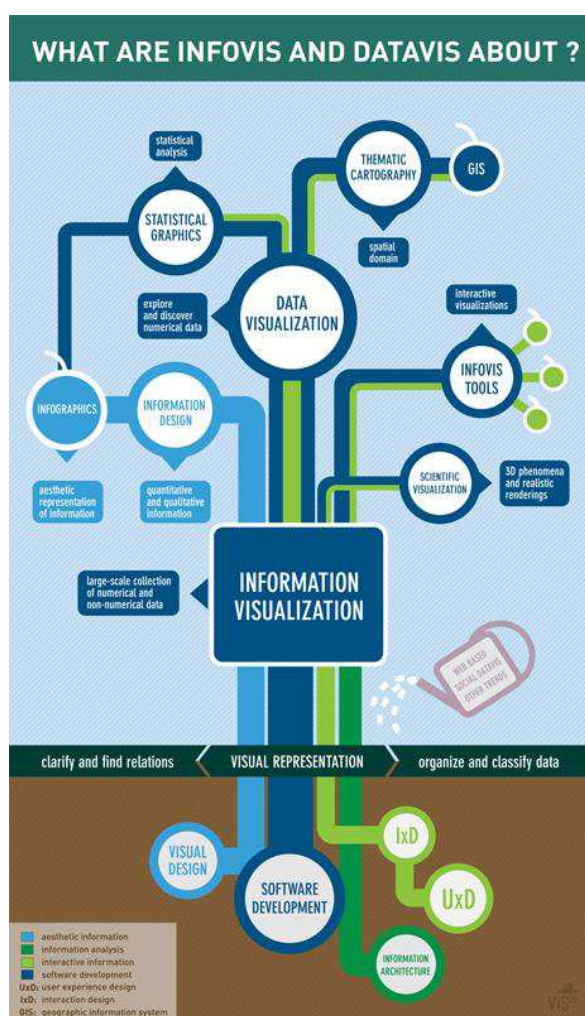
V současné době neexistují jasně definované hranice mezi vizuální analýzou a více tradiční vizualizací vědeckých dat či informací. Vizuální analýza se soustředí na smysluplný proces, který začíná získáváním dat, pokračuje opakováním a vylepšováním postupů, tak aby mohl uživatel na problém nahlédnout z různé

¹¹ V původním znění: „*spatial data visualization*“

perspektivy a testovat různé hypotézy. Proces končí získáním vhledu uživatele do problému. Soustředí se na zpracování příliš objemných nebo komplexních datasetů, které nemohou být pochopeny pomocí jednoduchého statického obrázku (Telea, 2015).

2.4 Přehled oblastí vizualizace

Jak již bylo zmíněno, termín vizualizace dat je obtížně definovatelný a nejsou stanoveny přesné rozdíly mezi jednotlivými pojmy v oblasti vizualizace. Následující informační grafika, pomocí užití metaforu stromu, zobrazuje oblast vizualizace dat a informací a znázorňuje, kolik různých oblastí do této kategorie spadá a jak jsou navzájem propojeny. Kořeny představují individuální dovednosti a schopnosti, které člověk potřebuje při tvorbě vizualizace. Z obrázku je patrné, že vizualizace je komplexní proces, ve kterém je obtížné zobrazovat a definovat jednotlivé její součásti jako rozdílné a striktní kategorie, které spolu navzájem nesouvisí (Rendgen et al., 2012).



Obrázek 4: Přehled oblastí vizualizace
zdroj: Rendgen et al. (2012), autor: Eloisa Paola Fontana

2.5 Cíle a účely vizualizace

Přestože na poli vizualizace rapidně vznikají stále nové techniky a nástroje, neměli bychom zapomínat na hlavní cíle a účely vizualizace. Na to poukazuje ve svém článku i Lorensen (2004). Uvádí, že vizualizace jako oblast počítačové vědy vznikla v reakci na potřebu vědeckých a biomedicínských komunit prozkoumat obrovské množství dat a získat náhled na fenomén představovaný těmito daty. Počítačová grafika v té době nesplňovala jejich vědecké požadavky, proto musela reagovat na potřeby své nové zákaznické základny. Lorensen uvádí, že pokud se v tomto oboru nezačneme soustředit na poskytování řešení na konkrétní problémy uživatelů, stane se z vizualizace pouze akademická disciplína, která má jen malý dopad v reálném světě.

Odpovědi na to, jakým směrem by se vývoj v této oblasti měl ubírat a jakou hodnotu má vizualizace, poskytuje článek van Wijka (2005), jehož hlavní body sumarizuje Telea (2015). Vizualizace musí být efektivní a účinná. Má dávat odpovědi na konkrétní otázky nebo pomáhat řešit problémy konkrétních uživatelů. Metodu považujeme za užitečnou, pokud používaná technika a nástroje vedou k takovým rozhodnutím, která prokazatelně sníží náklady a zvýší zisk.

Efektivnost vizualizace lze změřit. Musíme být nejprve schopni kvalitativně i kvantitativně popsat problém, který chceme vyřešit nebo otázku, na kterou chceme získat odpověď. Také musíme umět vyčíslit náklady na použití určité vizualizační metody. Náklady jsou buď technického rázu (rychlost procesoru, paměti počítače) nebo hůře vyčíslitelné (uživatelské úsilí získat vhled do problematiky). Nejlepší cesta, jakou můžeme měřit úspěch dané vizualizační techniky je počet uživatelů nebo vývojářů, kteří si tuto techniku osvojili. Proto je nutná spolupráce odborníků a koncových uživatelů při vývoji navrhovaných technik a nástrojů. Vývoj také nesmí probíhat v izolovaném prostředí, ale v integraci s prezentačními nebo průzkumnými metodami. Pravděpodobně největší výzvou v oblasti navrhování zdařilé vizualizace je zjištění, zda k ní stačí využít nebo přizpůsobit stávající nástroje a techniky nebo je potřeba vynalézt nové řešení na míru (Telea, 2015).

Záměrem vizualizace je získat pomocí interaktivní grafiky vhled do různých hledisek dané problematiky související s procesem, který nás zajímá, jakým je například vědecká simulace. S termínem vhled (insight) se setkáváme v textech o vizualizaci často. Efektivní

vizualizace nám poskytne odpovědi na konkrétní otázky ohledně dané problematiky a fakta, o kterých jsme nevěděli. Tímto způsobem získáváme vhled do daného problému (Telea, 2015).

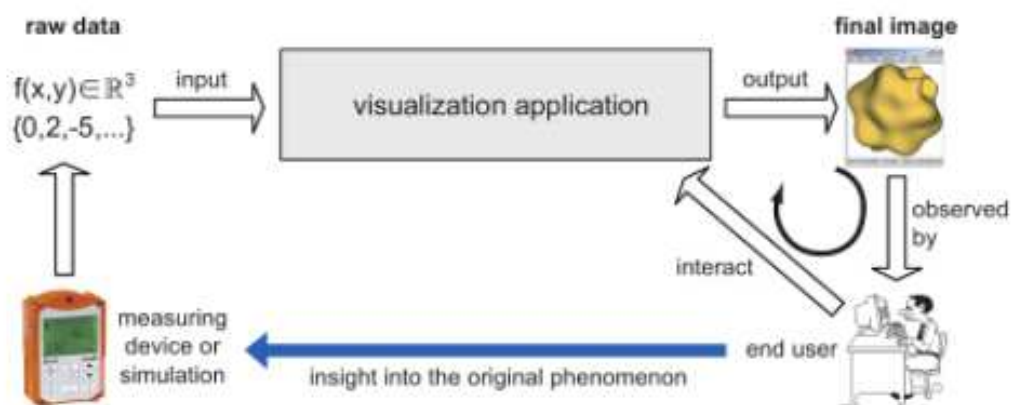
Vizualizace slouží ke sdělování příběhů, výzkumných poznatků a také jako platforma pro manipulaci a průzkum dat. Efektivní vizualizace nám pomáhají vytvářet znalosti. Vizualní zobrazení informací lze považovat za kognitivní artefakty, protože mohou posílit naše mentální schopnosti. Meirelles (2013) zkoumá vizualizace ve vztahu ke kognitivním principům, které jsou pro ně základem, což může být kombinace následujícího: zaznamenání informace, sdělení významu, zlepšení pracovní paměti, usnadnění vyhledávání a objevů, podpora vnímání, posílení detekce a rozpoznávání a poskytování modelů skutečného a teoretického světa.

3 Proces vizualizace

K vytvoření smysluplné vizualizace je zapotřebí kombinace uměleckých, statistických a matematických dovedností. Postupně se využití vizualizace dat stává stále populárnější metodou, jak co nejrychleji a neúčinněji sdělovat informace.

Tvoření vizualizací totiž neznamena pouze převedení tabulky dat do vizuální podoby. Vizualizace by měla komunikovat data co nejefektivněji, tedy skutečně odhalit data rychle a přesně. Vytvářením vizualizací můžeme snadno shrnout a sdělit data jiným lidem, což umožňuje pochopit i největší nebo nejsložitější sady dat (Laumans, 2009).

Jedním ze základních rysů vizualizace je její interaktivní aspekt. Pokud chceme vizualizovat velké množství dat, nemůžeme výsledný obraz bez dalšího zpracování přizpůsobit na jednu obrazovku. Protože existuje mnoho způsobů, jak vytvářet vizualizace ze surových dat, může si je uživatel sám vyzkoušet, aby více porozuměl celému procesu (Telea, 2015).



Obrázek 5: Konceptní pohled na proces vizualizace
zdroj: Telea (2015)

Z obrázku je patrné, že vstupním prvkem v procesu vizualizace jsou surová data pocházející z měřicího zařízení nebo simulace určitého procesu. Tato data se zpracovávají pomocí vizualizačního nástroje tak, že výsledkem je finální obraz vizualizace. Konečný uživatel může výsledek vizualizace nejen pouze pozorovat, ale pomocí vizualizačního nástroje na něj také působit a manipulovat s daty. Uživatel je

činitelem samotného procesu vizualizace tak, aby výsledkem bylo získání vhledu do původního jevu a pochopení zkoumaného fenoménu.

Vstupními prvky vizualizace jsou tedy data – subjekty, které samy o sobě nemají žádný význam. Jsou to pouze stavební kameny, pomocí kterých budeme vytvářet informace. Abychom jim dali význam, musíme data nejprve zpracovat, organizovat a prezentovat ve vhodných formátech. Když k informacím přidáme zkušenosti, získáme znalosti. Nejvyšší úroveň porozumění je v konečném důsledku moudrost. Můžeme ji definovat jako fázi, kdy člověk má takovou pokročilou úroveň znalostí o procesech a vztazích, že může podat kvalifikovaný úsudek o datech. Na rozdíl od vědomosti nemůže být přímo přenášena nebo vyučována (Mazza, 2009).

Dle Tufteho (2013, s. 13) spočívá dokonalost ve statistické grafice komplexností myšlenek, které jsou sdělovány jasně, precizně a efektivně. Výsledné grafické zobrazení by mělo *„ukázat data; přimět diváka, aby přemýšlel spíše o podstatě než o metodologii, grafickém designu, technologii grafické produkce nebo něčem jiném; vyvarovat se zkreslení toho, co by data měla ukazovat; učinit velké datasety konzistentní; ... sloužit jasnému účelu: popsat, prozkoumat, uspořádat nebo designovat data Grafika nám odhaluje data.“*¹²

Vizuálně atraktivní grafické zobrazení často nalézáme v jednoduchosti designu a komplexnosti dat. Vizualizace získává sílu z obsahu a interpretace dat místo pouhého okamžitého zobrazení nějakých čísel. Dle Tufteho (2013) by mělo působivé zobrazení dat mít správně zvolen formát a design, použít slova, čísla i obrazy, reflektovat rovnováhu, poměr a smysl pro relevantní měřítko, zobrazovat dostupnou komplexnost detailů, vyprávět pomocí dat příběh a vyvarovat se nesmyslné dekorace a „chartjunku“¹³.

Existuje mnoho způsobů, jakými můžeme vytvářet vizualizace dat. Způsoby se značně liší v závislosti na obsahu dat a účelu vizualizace (Laumans, 2009).

¹² V původním znění: *„show the data; induce the viewer to think about the substance rather than about methodology, graphic design, the technology of graphic production, or something else; avoid distorting what the data have to say; make large data sets coherent; ... serve a reasonably clear purpose: description, exploration, tabulation, or decoration ... Graphics reveal data.“*

¹³ nadbytečné prvky v grafu, které pro porozumění nejsou podstatné nebo naopak odpoutávají pozornost od hlavního sdělení; tento termín je podrobněji rozebrán v čtvrté kapitole

3.1 Efektivita vizualizace dat

Jestliže pracujeme s obrovským množstvím dat, je nesmírně složité získat komplexní pohled na jejich význam. Dalším problémem je také neustále se měnící povaha dat, což může vyústit v přidání nových informací k těm stávajícím, nebo přepracováním starších informací (Fry, 2008).

Zlepšujeme se ve shromažďování dat, ale zaostáváme v tom, co s nimi můžeme dělat. Jak na otázky najít odpovědi rychle a efektivně, je největší výzvou dnešní doby bohaté na informace. Fry (2008) si pokládá otázku proč jsme tak dobří v měření a zaznamenávání věcí, ale zůstáváme pozadu s metodami, jak těmto informacím porozumět a sdělovat je.

Máme tendenci považovat data za pevné, neměnné hodnoty, musíme si ovšem uvědomit, že data se neustále mění, protože pocházejí z reálného světa, kde nic není absolutní. V dnešní době již nemáme problém s nedostatkem místa, kde uchovávat data. Proto je snazší oddělit data od původního účelu jejich shromažďování. Pokládáme si však otázku, jak můžeme pochopit tak obrovské množství dat (Fry, 2008).

Každý soubor dat můžeme vizualizovat několika způsoby, které jsou různě efektivní. Abychom mohli vytvořit efektivní vizualizaci musíme data nejdříve pochopit. Je podstatné nalézt odpovědi na otázky: co jsou to za data, jaké jsou vztahy mezi proměnnými, jak jsou data organizována a zejména co bychom ve výsledku chtěli uživateli sdělit. Odpovědi na tyto otázky budou určovat formu vizualizace (Laumans, 2009).

Při každé tvorbě vizualizace dat se její autor zaměřuje na něco jiného. Základní zaměření by ale mělo být vždy stejné – hledáme vzory a souvislosti mezi daty. Zkoumáme data, jak se mění v průběhu času, jaká událost vyvolává změnu, kdy se situace opakuje a podobně. Hledání souvislostí tvoří přidanou hodnotu efektivní vizualizace než jen pouhé náhodné umístění grafů. Ve statistice to obvykle znamená hledání korelací a příčinné souvislosti. O data sítích bychom neměli přemýšlet jako o oddělených skupinách, ale jako o několika oddílech, které jsou spolu ve vzájemné interakci. Měli bychom si vždy položit otázku, co vlastně vidíme a důkladně data zkontrolovat a verifikovat (Yau, 2011).

Aby vizualizace dat měla smysluplný výsledek, musíme si v rámci procesu osvojit několik různorodých disciplín, jako například statistiku, dolování dat, grafický design a další. Tyto disciplíny jsou od sebe často izolovány. V rámci grafického designu se nezabýváme tím, kde vezmeme data a podobně u dolování dat se nezabýváme tím, jak bude vypadat jejich následná vizualizace (Fry, 2008).

Fry (2008) tvrdí, že musíme sladit jednotlivé disciplíny tak, aby byly součástí jednoho procesu. Měli bychom se soustředit na to, jak jsou data chápána komplexně spíše než na hledisko každého jednotlivého procesu a nástrojů, které používáme. Proces porozumění datům začíná souborem čísel a otázkou, k odpovědi vedou následující kroky:

získání dat, strukturace (zpracování) dat a zařazení do jednotlivých kategorií, vyfiltrování nepotřebných údajů, vytěžení dat, zobrazení dat v základním vizuálním modelu (sloupcový graf, seznam nebo stromová struktura), vylepšení vizualizace, tak aby byla jasnější a působivější a konečně přidání uživatelské interaktivity jako metody pro manipulaci s daty nebo přidání ovládacích prvků, které skryjí/zobrazí jednotlivé funkce (Fry, 2008; Yau, 2011).

3.2 Základní schémata vizualizace dat

Při tvorbě vizualizace dnes obvykle již používáme předem připravená základní schémata pro vizualizaci. Dle Marka (2014, s. 22) jsou nynější běžné statistické grafy „*ustálené metaforou informační vizualizace, ... které jsou ověřeným způsobem grafické reprezentace kvantitativních dat.*“

Nejběžnější formou vizualizace dat jsou jednoduché **sloupcové grafy**, které srovnávají hodnoty nezávislých proměnných.

Spojnicový graf používáme pro kontinuální data, kde horizontální osa obvykle představuje časové období. Tento typ grafu ukazuje, jak se proměnná vyvíjí v průběhu času.

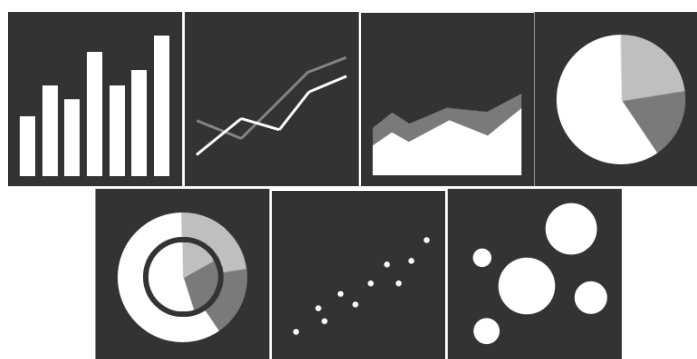
Plošný graf je podobný spojnicovému grafu, ale vykreslená oblast je součtem všech proměnných, pocházejících z různých oblastí.

Výsečový graf je nejběžnějším typem grafu, který je využíván pro rychlé srovnání poměrných dat. Používá se, když data představují části celku.

Prstencový graf podobně jako výsečový zobrazuje části celku, avšak může porovnávat více než jednu sadu dat, kdy každý prstenec znázorňuje právě jednu datovou řadu.

Bodový graf zobrazuje hodnoty dvou proměnných. Jsou často používány k nalezení vztahů mezi daty nebo k odhalení informací, které nejsou viditelné v prosté tabulce.

Bublinový graf zobrazuje navíc hodnotu ještě třetí proměnné pomocí velikosti samotné bubliny (Laumans, 2009; Marek, 2014).



Obrázek 6: Základní typy grafů
zdroj: Laumans (2009)

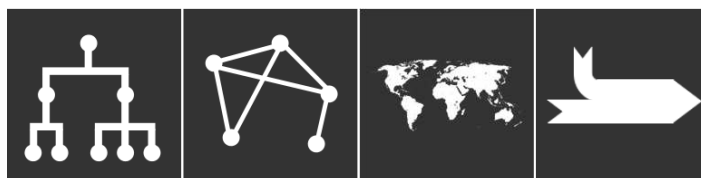
3.3 Pokročilé grafy

Stromový diagram se používá při reprezentaci hierarchie dat, například u vizualizace rodokmenu.

Síťové vizualizace se používají v případě, kdy je nejdůležitějším prvkem vizualizace ukázat, která data jsou navzájem propojena, na rozdíl od toho jakým způsobem.

Kartografická mapa slouží pro data, která jsou relevantní pro konkrétní místa nebo oblasti, které lze vykreslit na mapě.

Sankey diagram použijeme, jestliže jsou data součástí procesu. Lze je poté zobrazit pomocí diagramu toku. Skládá se z několika menších kanálů, které se spojují do jednoho hlavního.



Obrázek 7: Pokročilé typy grafů
zdroj: Laumans (2009)

Výše jsou uvedeny pouze některé z nejpobulárnějších typů vizualizací. Vizualizace jsou dle stávajících vzorů efektivní, protože nejen že účinně pracují pro určitý typ dat, ale také jsou s nimi lidé obeznámeni, a to jim usnadňuje jejich pochopení. Můžeme se však setkat s daty, pro které musíme hledat nové způsoby vizualizace, popřípadě vytvořit kombinaci několika vzorů (Laumans, 2009).

V minulosti byly všechny vizualizace statické a předdefinované, uživatel nemohl provádět jakýkoli typ interakce. Z tohoto důvodu vizualizace vytvářela pouze jeden nezměnitelný pohled. Dnes vlivem rozvoje informačních technologií může uživatel manipulovat s procesem, který vytváří konečnou grafickou reprezentaci pomocí přiblížení, rotace, posunu nebo prostřednictvím filtrování dat přímo ovlivnit a upravit generované vizualizace (Mazza, 2009).

4 Aktuální trendy v oblasti prezentace dat

4.1 Storytelling

Jakým způsobem lze prezentovat a komunikovat data bylo dlouhou dobu na okraji zájmu v oblasti výzkumu vizualizace, jelikož byl kladen důraz především na průzkum a analýzu samotných dat. Kosara a Mackinlay (2013) se však domnívají, že prezentace dat za pomoci vyprávění příběhů (dále jen *storytelling*) je stejně důležitá, jako ostatní oblasti vizualizace dat a měl by se na toto téma zaměřit výzkum. Storytelling dá informacím strukturu, která je snadno zapamatovatelná, což je důležité zejména v případě, kdy běžně není ten, co analyzuje data stejnou osobou, jako ten, který dělá rozhodnutí nebo jen potřebuje informace sdílet s ostatními.

V raných fázích vizualizace se výzkum zabýval průzkumem a analýzou dat, a proto byly vynalezeny techniky pro většinu data setů a úloh. Techniky, jak prezentovat a komunikovat data, stále schází. Ale právě spojení faktů do příběhů je jedním z nejefektivnějších způsobů, jak data prezentovat a argumentovat jimi. Příběhy jsou nejen populární způsob, jak zachovat informace a předávat je, ale také spojují fakta tak, aby byly zapamatovatelné. Příběh si můžeme definovat jako uspořádaný postup jednotlivých kroků, který má svůj začátek a konec. V oblasti vizualizace jsou to příběhy založené na datech a obvykle se skládají z textů a obrázků, popř. videí. Klíčovým rysem příběhů je řád, čas je důležitý pro pochopení kauzality – události dřívější mohou ovlivnit ty pozdější, ale nikdy ne naopak. Výzkumníci se často mylně domnívají, že nástroje, které používají pro analýzu dat budou také vhodné pro jejich prezentaci. S tímto názorem však nelze souhlasit tvrdí Kosara a Mackinley (2013).

Jak již bylo zmíněno v kapitole o historii vizualizace, některé vizualizace (například mapa Johna Snowa či Charlese J. Minarda) sloužily již tehdy primárně k vyprávění příběhu než jako analytický nástroj pro porozumění komplexním datům.

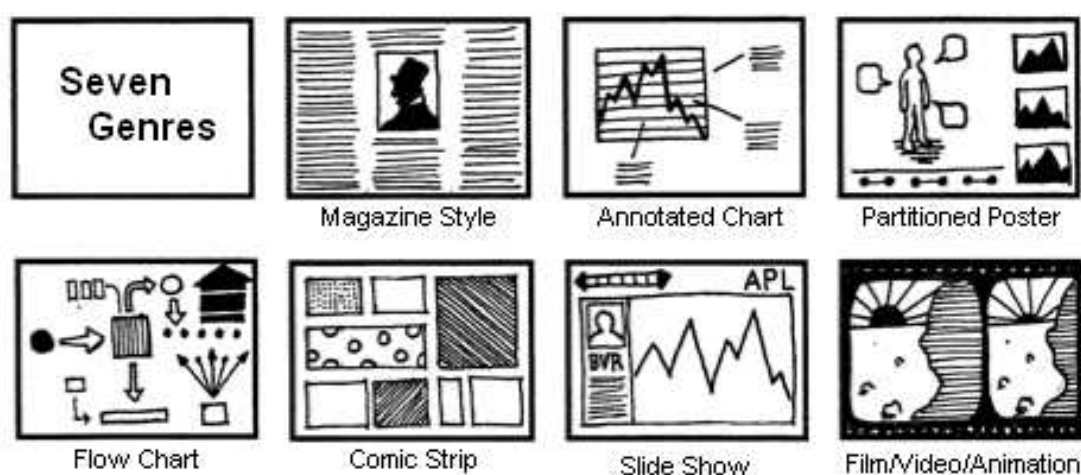
Gershon a Page (2001) uvádí, že storytelling pomůže vizualizaci odhalit informace tak účinně a intuitivně, jako bychom sledovali film. Již od dob vynálezu tisku se stále vyvíjejí nové a sofistikované metody, kterými můžeme sdělovat příběhy. Kvalitní storytelling sděluje velké množství informací ve vhodném formátu a pomocí relativně málo slov. Příběhy jsou více působivé a informace v nich lépe chápány, než když jsou nám informace předány například pomocí jednoduchých seznamů.

Pomocí obrázků může uživatel informace pochopit také poměrně rychle. Vizualizací ale nelze vyjádřit vše, a proto může docházet k různým nepochopením toho, co daná vizualizace znázorňuje. Případné nepochopení vyjasní právě doplňující storytelling, proto je tato kombinace vysoce efektivní (Gershon a Page, 2001).

Storytelling je pokládána za cennou součást vizualizace informací. Informace zpravidla přichází v reálném čase a uživatel je musí zpracovat, pochopit a vyhodnotit. Gershon a Page (2001) srovnávají vizualizaci s médii tradičního zábavního průmyslu, kdy obě formy usilují o udržení divákovy pozornosti při sdělování informací. Rozdíl se ale nachází v komplikovanosti sdělovaných informací, kdy prostřednictvím vizualizace chceme komunikovat mnohem složitější informace než například při sledování filmu.

Abychom mohli efektivně prezentovat informace s použitím metody storytellingu, musíme mít nejen vlastnosti technických expertů, ale také filmových režisérů, abychom diváka uměli nejen zaujmout, ale i prezentovat relevantní data (Gershon a Page, 2001).

Segel a Heer (2010) ve své práci analyzovali 58 příkladů narativních vizualizací¹⁴, na jejichž základu poté popsali sedm stylů, kterými můžeme komunikovat data pomocí vyprávění – časopisecký styl, anotovaný graf, rozdělený plakát, vývojový diagram, komiks, prezentace a video.



Obrázek 8: Styly vyprávění
zdroj: Segel a Heer (2010)

¹⁴ V původním znění: „*narrative visualization*“

Tyto styly se navzájem nevylučují, naopak jejich kombinací vznikají ještě komplexnější vizualizace. Všechny styly jsou používány k vyprávění příběhů, ale některé fungují na daný typ příběhu efektivněji než jiné. Je důležité si zvolit odpovídající styl na základě různých faktorů – jaká data a příběh chceme sdělit, jaké je naše publikum a pomocí jakého média budeme příběh sdělovat. Důležitými prvky v tomto typu vizualizace jsou textové prvky a interaktivita. Texty jsou využívány pro vysvětlení obrázků (nadpisy, anotace). Interaktivita dovoluje uživateli manipulovat s vizualizací pomocí navigačních tlačítek, detailů na vyžádání, filtrování, hledání, přibližování a dalších možností. Textové prvky mohou pomoci vyjasnit prvky vizuální, v případě nesprávného použití ale vytvářejí chaos. Interaktivita na jednu stranu může angažovat uživatele, ale také hrozí vzdálení se od autorovy zamyšlené zprávy (Segel a Heer, 2010).

Ve studii dále uvádějí dva typy přístupů – autorský a čtenářský¹⁵. V prvním případě se jedná o lineární průběh s množstvím textových prvků a žádnou interakcí (například video). Ve druhém případě neexistuje předepsané přesné pořadí obrázků, nenachází se zde žádné textové prvky, ale pracuje se s vysokou mírou interaktivit¹⁶. Většina vizualizací se pohybuje mezi těmito dvěma přístupy. Na závěr autoři popsali tři schémata: Martini Glass structure, interaktivní prezentace a Drill-Down story (Segel a Heer, 2010).

Martini Glass Structure¹⁷ upřednostňuje autorský přístup, po kterém se otevře prostor pro čtenáře, kde může uživatel volně zkoumat data (Kosara a Mackinlay, 2013).

Interaktivní prezentace se snaží najít balanc mezi dvěma přístupy. Tato struktura je navržena tak, aby mohl uživatel prozkoumat konkrétní body ještě předtím, než se prezentace dostane do další fáze. Zajišťuje, aby uživatel následoval další kroky v prezentaci, jakmile bude spokojen s průzkumem dat (Segel a Heer, 2010).

Drill-Down Story¹⁸ prezentuje základní téma a nechává uživatele, aby si sám vybral jednotlivé případy, které ho zajímají a odhalil tak další informace. Tato struktura nechává uživatele diktovat, jakým způsobem se bude příběh dále vyvíjet, ale stále vyžaduje

¹⁵ V původním znění: „*author-driven, reader-driven*“

¹⁶ například pomocí vizualizačního nástroje Tableau, který je blíže popsán v páté kapitole a v praktické části

¹⁷ jedná se o trojúhelníkový tvar sklenice na stopce, která zobrazuje lineární průběh a konečné rozšíření značí otevření se interakci

¹⁸ podívat se, prozkoumat něco do hloubky

i přítomnost autora vizualizace, který určuje možné typy interakce, jaké příběhy budou komunikovány a jaké detaily budou sděleny (Segel a Heer, 2010).

Na základě analýzy vybraných příkladů došli autoři Segel a Heer (2010) k závěru, že prezentace dat s využíváním storytellingu je nejvíce efektivní, jestliže existují pouze omezené interakce v různých kontrolních bodech vyprávění, pomocí kterých je dovoleno uživatelům zkoumat data bez uhýbání od autorova zamýšleného tématu.

4.1.1 Specifika storytellingu u vědeckých dat

V předchozí kapitole byl kladen důraz zejména na storytelling v oblasti vizualizace informací. Jak je to ale se storytellingem pokud chceme použít vědecká data? Tímto fenoménem se zabývá Kwan-Liu Ma et al. (2012). Poukazuje na to, že vyprávění příběhů s pomocí vědeckých dat je něco, čemu se věnuje málo pozornosti, přestože formování dat do příběhu je pro diváka, jak už bylo zmíněno, více zajímavé a také snadněji zapamatovatelné. Narativní dopad na vizualizaci vědeckých dat je takový, že jsme schopni vidět neviděné – vizualizace funguje jako rozšíření našich smyslů. Vizualizaci takovýchto přesných dat tvoří vědci zejména pro ostatní vědce, proto je tvorba dané vizualizace v určitých ohledech jednodušší – téma se nemusí určenému publiku široce představovat. Problém nastává v okamžiku, pokud bychom výsledky dané vizualizace chtěli předat i široké veřejnosti (Ma et al., 2012).

Rostoucí význam vizualizací představuje příležitost například pro muzea. Vizualizace vědeckých dat jim může poskytovat takové obrázky, které se setkají se zájmem publika, a to se tak seznámí s fenoménem, o němž nikdy předtím neslyšeli. Klíčová je ovšem správná interpretace dat a vysvětlení tohoto fenoménu (Ma et al., 2012).

Kwan-Liu Ma (2012) shrnuje, že daná vzrůstající role vizualizací ve vědecké sféře představuje pro muzea a podobné instituce možnost, jak zapojit širokou veřejnost. Musíme si však uvědomit, že přeměna zvyšujícího počtu vizualizací vědeckých dat na smysluplný zážitek pro širokou veřejnost vyžaduje správnou interpretaci, design a spolupráci. Takováto vizualizace dat s vyprávěním zahrnuje srozumitelnost, důvěryhodnost a angažovanost publika. Pokud bude uživatel aktivně participovat na procesu vizualizace, pocítí smysl pro angažovanost s daty, které jsou prezentovány. Je třeba zvážit, jak storytelling a vizualizace mohou vědecké poznatky učinit více srozumitelnými a přístupnými pro širokou veřejnost. Inspiraci jí může poskytnout proces

vizualizace informací. Důležité je znát své publikum, aby se vizualizace stala přístupnější. Klíčové je správné nastavení spolupráce vědců a designérů.

Takovým příkladem spolupráce je Národní muzeum americké historie (*Smithsonian Institution's National Museum of American History*), které se inspirovalo prací Bena Shneidermana, vynálezce stromových map (proporční graf). Muzeum založilo webovou stránku *HistoryWired: A Few of our Favorite Things*, která byla přístupná mezi lety 2001–2016. Byla navržena tak, aby mohl uživatel prozkoumat celkem 450 objektů muzejní sbírky se zajímavými příběhy k vyprávění¹⁹. Objekty byly rozděleny do různých obecných kategorií, detaily k nim bylo možno zobrazit kliknutím na jednotlivé čtverce. Hlavními zásadami tohoto designu byla customizace²⁰, přímá manipulace, zpětná vazba, textové vyhledávání a další interaktivní prvky (Bailey a Pregill, 2014).

4.2 Informační grafika

V literatuře nalezneme množství definic pojmu informační grafika (zkráceně infografika). Smiciklas (2012, s. 17) definuje informační grafiku jako „vizualizaci dat nebo myšlenek, které se snaží sdělit komplexní informace publiku takovým způsobem, který lze rychle konzumovat a pochopit“²¹. Poněkud stručnější je Nigel Holmes, který infografiku označuje jako „vysvětlující grafiku“²².

Informační grafika znamená více než pouhou vizuální reprezentaci dat. „Dnes ... infografiky ... kombinují vizualizaci dat, ilustraci, text a obrázky do smysluplného výstupu, který nám poskytuje úplný příběh“²³ (Krum, 2014, s.6). Vizualizace dat je tedy jen jedním ze samostatných prvků, který je v infografice použit. Jejím cílem je informovat, pobavit nebo přesvědčit publikum. Většina infografik se skládá ze tří částí – úvodu, klíčového sdělení a závěru. Úvod slouží k získání čtenářovy pozornosti a obsahuje základní informace, které čtenář potřebuje vědět a pochopit, než se pustí do detailního zkoumání. Klíčové sdělení by mělo obsahovat pro čtenáře dosud

¹⁹ Více informací o projektu: <http://americanhistory.si.edu/exhibitions/history-wired>

²⁰ přizpůsobení individuálním potřebám zákazníka

²¹ V původním znění: „a visualization of data or ideas that tries to convey complex information to an audience in a manner that can be quickly consumed and easily understood“

²² V původním znění: „explanation graphics“

²³ V původním znění: „Today....inphographics ... combines data visualizations, illustrations, text, and images together into a format that tells a complete story.“

neznámé informace a mělo by jej zaujmout natolik, že si tento hlavní bod informací zapamatuje nejvíce.

Ve spodní části by měla mít grafika závěr, popř. výzvu k akci, jestliže existuje nějaký postup, který by měl čtenář následovat poté, co se dozvěděl danou informaci (například koupit si produkt, navštívit webovou stránku, podepsat petici apod.) (Krum, 2014).

4.3 Vliv vizualizace na paměť

Jak již bylo uvedeno, cílem prezentace dat je nejen zaujmout diváka, ale také snaha o to, aby pochopil a zapamatoval si dané téma. Přestože je efekt vizualizace pro paměť důležitý, není tato oblast ještě prozkoumána do hloubky (Kosara a Mackinlay, 2013).

Zajímavým počinem v této oblasti je práce Batemana et al. (2010), který ve svém výzkumu zkoumal, zda má „junk“²⁴ vliv na pochopení a zapamatování si. Vycházel z poznatku, že v některých návodech, jakým způsobem se mají tvořit informační grafy, nalezneme doporučení, aby se prezentace vyvarovala „*chart junku*“ tedy všech zdobných prvků, které nejsou podstatné pro pochopení dat. Kritikem těchto vizuálních ozdob²⁵ je například Edward Tufte, který tvrdí, že tyto zdobné prvky odvádí čtenáře od dat a mohou způsobit jejich dezinterpretaci (Bateman et al., 2010, Tufte, 2013). Tufte, nazýván „*the da Vinci of data*“ (Shapley, 1998), je jedním z nejvlivnějších odborníků v oblasti vizualizací a informačního designu. Zabývá se nejen historií statistické grafiky, ale položil základy oboru vizualizace informací (Bailey a Pregill, 2014).

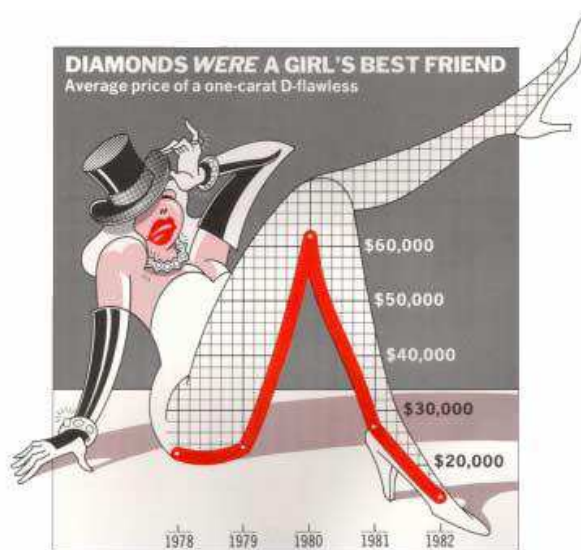
Opačný názor na „*chart junk*“ má Nigel Holmes, tvrdí, že datová grafika musí čtenáře zaujmout a něčím se odlišit od ostatních grafů (Bateman et al., 2010). Holmes během své práce v časopise Time vytvářel „*vysvětlující grafiky*“, které doprovázely články. Pomocí humoru a emocí prezentuje data nezapomenutelným a vizuálně atraktivním způsobem, který čtenáře zaujme. Pro Tufteho znamená takovýto „*chart junk*“ pohrdání informacemi i publikem. Debata, zda „*chart junk*“ brání či pomáhá pochopení informací dále pokračuje (Bailey a Pregill, 2014).

Bateman et al. (2010) proto v experimentu zkoumal přesnost výkladu grafu a zapamatování si jej. Předložil účastníkům výzkumu dva grafy obsahující stejné sdělení

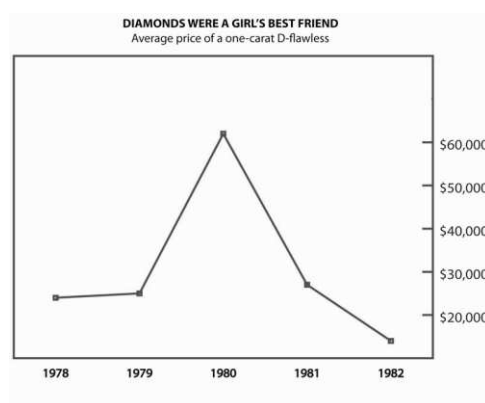
²⁴ přílišné zdobené, nepodstatné/nemající vliv na sdělování faktů

²⁵ V původním znění: „*visual embellishment*“

a data – jeden čistý a druhý se silným grafickým zdobením („Holmes-style graf“). Po dvou týdnech si měli účastníci výzkumu vybavit co nejvíce detailů. Z experimentu bylo zjištěno, že zdobnost/čistota grafu nemá vliv na přesnou interpretaci dat, po delší době si ale lidé mnohem více vybavují detaily zdobených grafů a klíčové sdělení v Holmesových grafech než v grafech bez grafické úpravy. Holmesovy grafy jsou pro uživatele také více atraktivní, snadnější na pochopení a rychleji zapamatovatelné. Bateman et al. (2010) ve svém výzkumu pracoval pouze s grafy bez jakékoli možné čtenářovy interakce.



Obrázek 9: Grafika od Nigela Holmesa
zdroj: Bateman et al. (2010)



Obrázek 10: čistý graf použitý v experimentu
zdroj: Bateman et al. (2010)

4.3.1 Vliv interakce na paměť

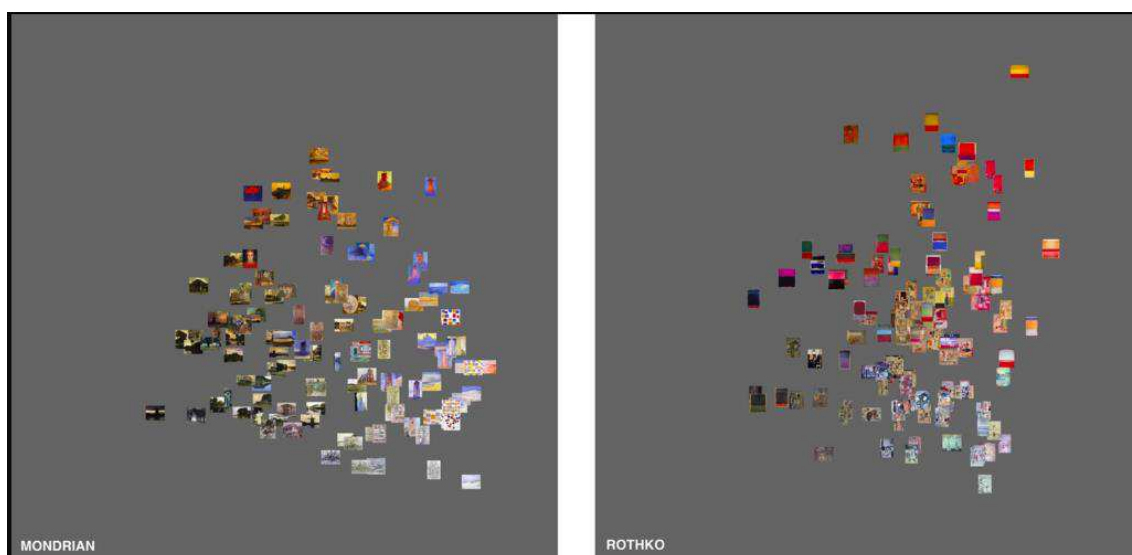
Dalším prvkem pro snadnější zapamatování si může být právě interakce – být schopný nejen vidět data, ale také s nimi pracovat, zobrazovat si jen určité prvky apod.

Příběhy jsou nutně spojeny s účastí publika a přirozeně vedou k otázkám, což vede k diskuzi a dále k hlubší analýze. Příběhy, zahrnující informace a znalosti, mohou být jednoduše šířeny další osobou. Dle Kosara a Mackinlaye (2013) je možné, že vizuální datové příběhy budou právě tím správným způsobem, jak zachovávat informace o komplexních datech a procesech a předávat je dál.

4.4 Přímá vizualizace

Pokroky v oblasti počítačové grafiky, softwarových nástrojů a digitální vizualizace vedly na konci dvacátého století k vizualizaci dat kulturního dědictví. Lev Manovich (2009) ustanovil nové paradigma – kulturní analytika. Dnešní vědci, agentury, vlády a soukromé podniky vizualizují velké datové soubory a spoléhají se na počítačem vytvořené analýzy. Používají statistickou analýzu dat, dolování dat, vizualizaci informací, vědeckou vizualizaci, vizuální analýzu a simulaci. Manovich navrhuje začít používat tyto techniky také pro kulturní data, která máme k dispozici.

Manovich používá přímou vizualizační metodu, u které se vizuální symboly pro datové objekty nenahrazují, ale jsou zobrazovány v původní podobě ve zmenšené velikosti.²⁶ Je tedy rozdílnou metodou od tradiční vizualizace informací, u které se zobrazují graficky primitivní prvky (jednoduché grafy, ikony) pro označení objektů a prokazování vztahů mezi nimi, (Bailey a Pregill, 2014). Příklad takovéto přímé vizualizace je uveden níže.



Obrázek 11: Modrian vs. Rothko; Lev Manovich (2009)

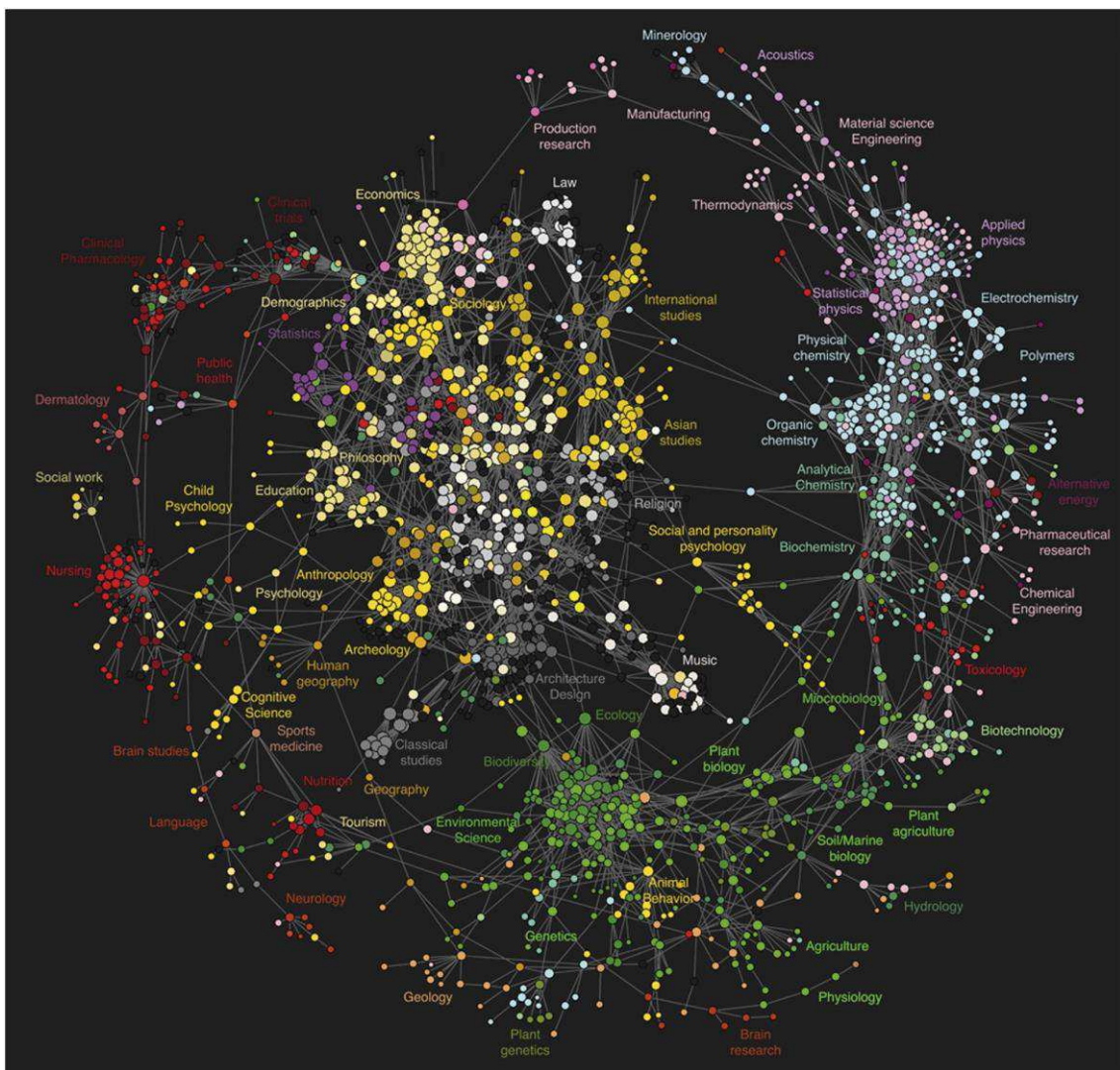
zdroj: <http://lab.softwarestudies.com/2011/06/mondrian-vs-rothko-footprints-and.html>

²⁶ srovnání s Kosarou (s. 17), vizualizace musí být založena na nevizuálních datech, v opačném případě se jedná obrazovou produkci nikoli o vizualizaci dat

Obrázek porovnává vizualizaci 128 obrazů Modriana a 123 obrazů Rothka, které byly malíři vytvořeny ve srovnatelném časovém období. Osa X označuje jas, osa Y saturaci. Vizualizace byla vytvořena pomocí softwaru ImagePlot Lvem Manovichem. Z této vizualizace je patrné, že v případě obou umělců se jejich první malby inspirovaly předchůdci a ostatními současníky, postupně si umělci vytvářeli svůj vlastní vizuální styl. Na konci sledovaného období se již jedná o abstraktní malby. Také si můžeme všimnout, že žádné dva obrazy se nenacházejí ve stejném prostoru na ose X nebo Y, tudíž jsou všechny obrazy dostatečně navzájem odlišné.

4.5 Síťové vizualizace

Posledním aktuálním trendem, kterému se budeme v této práci věnovat, jsou síťové vizualizace. Tento typ vizualizací hledá vzorce a smysluplné spojení mezi jednotlivými prvky. Manuel Lima (2011) zkoumá, jak si lidé organizují a vizualizují informace v průběhu věků. Uvádí posun od metafory stromu k metafoře sítě. V minulosti byly informace strukturované a organizované. Stromový diagram se využíval k zobrazení vědění, morálky, právního systému apod. Nyní lidstvo produkuje tolik informací, že již není možné vše zpracovat do hierarchické podoby. Sítě nejsou pouze jedním z trendů vizualizace, usilují o mapování různorodosti složitých systémů. Zaměřují se na decentralizaci, provázanost mezi jednotlivými prvky a vzájemné závislosti. Lima uvádí, že tento nový způsob myšlení je zásadní při řešení složitých problémů dnešní doby. Příkladem síťové vizualizace je mapa vědních disciplín Bollena et al. (2009), kde kruhy zobrazují jednotlivé odborné časopisy a spojující čáry označují uživatelské chování při procházení jednotlivých webových stránek. Výsledná vizualizace zobrazuje vztahy mezi různými vědeckými obory a ukazuje propojení společenských a humanitních věd s přírodními vědami.



Obrázek 12: Mapa vědních disciplín J. Bollen, H. Van de Sompel, A. Hagberg, L. Bettencourt, R. Chute, et al. (2009)
 zdroj: <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0004803>

5 Nástroje k vizualizaci dat

V dnešní době se většina vizualizací dat provádí prostřednictvím počítače. Vizualizace pomocí počítačových softwarových nástrojů jsou logickou volbou z několika důvodů – surová data máme k dispozici v elektronické podobě, jedná se o velké objemy dat, výsledná vizualizace nemusí mít konečnou podobu (lze měnit výslednou podobu i vstupní data), výsledek se dá také snadněji sdílet a je jednoduše dostupný online.

Telea (2015) ve své knize poskytuje přehled několika užitečných nástrojů pro vizualizaci různých typů dat a pojednává o klasifikaci softwarových nástrojů pro vizualizaci. Dle něj musí být takovýto software efektivní (vytvořit vizualizaci rychle), škálovatelný (umí si poradit s velkými objemy dat), snadně se používat, být přizpůsobitelný požadavkům uživatele a snadno dostupný.

Nástroje pro vizualizaci můžeme rozdělit dle několika kategorií:

- placené nebo zdarma,
- open source nebo komerční,
- (ne)vyžadující znalosti programování,
- podle použití k vizualizaci vědeckých dat nebo vizualizaci informací,
- podle třídy:
 - aplikační rámec (application framework) – jedná se o sadu nástrojů, které mají za cíl ušetřit práci při programování pomocí obecných modulů, které mohou být opakovaně využity, namísto programování stále stejných prvků,
 - na klíč (turnkey systems) – systém, který je přizpůsoben pro konkrétní aplikaci, disponuje veškerým hardwarem a softwarem umožňujícím začít pracovat okamžitě,
 - knihovna – používá API (aplikační rozhraní) a umožňuje použití již vytvořených kódů v jiných programech.

Pro vizualizaci vědeckých dat lze použít například tyto softwarové nástroje: The Visualization Toolkit (VTK), MeVisLab, AVS/Express, SCIRun nebo ParaView.

Pro vizualizaci informací jsou vhodné například tyto nástroje: The InfoVis Toolkit (IVTK), GraphViz, Tulip, Gephi, ManyEyes nebo Treemap.

5.1 Out-of-the-box software²⁷

Tento typ softwaru je nejjednodušší k použití, a tedy vhodný i pro začátečníky. Nejznámějším je *Microsoft Excel* či cloudová alternativa v podobě *Google Spreadsheets*, pomocí kterých lze vytvořit všechny typy základních grafů. Dalšími představiteli tohoto typu softwaru jsou například:

Many Eyes – projekt společnosti IBM a jejich Visual Communication laboratoře. Jedná se o online aplikaci, pomocí které můžeme tvořit interaktivní vizualizace nabízející mnoho možností. Vytvořené vizualizace jsou dostupné online, lze je tedy snadno sdílet a jsou dostupné i všem ostatním uživatelům – odtud pramení i název aplikace. Nevýhoda spočívá právě v dostupnosti, kdy vše, co nahrajeme, je veřejně dostupné, tudíž nemůžeme vizualizovat citlivá data.

Tableau Software – v poslední době velmi oblíbený software, má příjemné uživatelské rozhraní (drag-and-drop systém²⁸, nemusíme mít technické znalosti), umožňuje zkoumat a vizualizovat data. U bezplatná verze podobně jako u *Many Eyes* budou vizualizace veřejně dostupné (Yau, 2011).

5.2 Programování

Některé programovací jazyky jsou pro vizualizaci vhodnější než jiné. Mezi nejběžnější patří *Python*, který nejen, že dokáže zpracovávat velké objemy dat, ale také umí vytvořit vizualizaci (nicméně není zcela esteticky působivá).

Dalším programovacím jazykem vhodným k práci s daty je *R*. Tento jazyk byl speciálně navržen pro analýzu dat, zejména pro statistické výzkumníky. Pro *R* je dostupných mnoho

²⁷ software umožňující okamžité použití, stačí jej pouze nainstalovat nebo spustit, není potřeba žádná konfigurace

²⁸ „táhni a pusť“ – uživatel uchopí daný objekt počítačovou myší a přesune ho přetažením na jiné místo, až poté tlačítko pusť.

balíčků (knihoven), takže je tvoření vizualizací poměrně jednoduché a rychlé. Na druhou stranu *R* není vhodným nástrojem pro interaktivní grafiku a animaci.

V *Pythonu* či *R* nejsou výsledné vizualizace vhodné ke konečné prezentaci, jelikož nejsou pro publikum dostatečně esteticky atraktivní. Proto se často následně upravují v grafických editorech (např. Adobe Illustrator, Inkscape). Výsledná vizualizace je poté nejen esteticky působivější, ale také čitelnější a jasnější, což by bylo mnohem obtížnější dosáhnout pouze s automaticky generovanými výsledky. (Yau, 2011).

Mezi další jazyky vhodné pro vizualizaci patří například Processing, PHP, Flash a Actionscript (pro interaktivní a animovanou webovou grafiku), dále HTML, JavaScript a CSS pro vizualizace, které fungují nativně ve webovém prohlížeči. Pro JavaScript také existují užitečné knihovny – jQuery, Protovis, InfoVis Toolkit (Fry, 2008, Yau, 2011).

V praktické části je předvedena vizualizace dat za pomoci některých z výše uvedených nástrojů a několika dalších vhodných k vizualizaci.

6 Praktická část procesu vizualizace dat

S vizualizací dat jsou spojeny mnohé překážky, které nejsou dodnes překonány. Jak již bylo zmíněno, v procesu shromažďování dat existují efektivní nástroje a technické problémy na této straně jsou méně časté. Jsme schopni shromažďovat velké množství dat efektivně a rychle. Zaostáváme ale v následné práci s takovým obrovským množstvím dat a tím, co s nimi lze vytvářet, abychom jim dali smysl. Fry (2008) poznamenává, že lidé jsou velice schopní v oblasti měření a zaznamenávání věcí, ale neudrželi krok i v oblasti metod, jak těmto informacím porozumět a dále je sdělovat. Fry představuje sedmistupňový proces návrhu vizualizace dat, který sladí všechny etapy do jednoho procesu:

- **získání dat** – z lokálního souboru nebo ze sítě,
- **strukturování dat** – vytvoření struktury a uspořádání do kategorií,
- **filtrování** – vyvarování se informačního přehlcení; odstranění všech dat, kromě těch, která nás zajímají,
- **těžení dat** – použití statistických metod nebo těžení dat k rozpoznání vzorů; uvedení dat do matematického kontextu,
- **reprezentace dat** – vytvoření základního vizuálního modelu (sloupcový graf, seznam),
- **vylepšení** – přetvoření základního zobrazení tak, aby bylo jasnější, čitelnější a vizuálně působivější,
- **interakce** – přidání metod pro manipulaci s daty, ovládání toho, které funkce jsou viditelné.

6.1 Kde získáme data?

Mít správná data je základem celého vizualizačního procesu. Ať už jsou data získána jakýmkoliv způsobem, vždy je nutné se zaměřit na to odkud jsou získána, jak byla sbírána, čeho se týkají a zda v nich nejsou chyby.

Data lze získat třemi způsoby:

- vlastní shromáždění dat,

- poskytnutí někým jiným – např. odborníky v oblasti, která nás zajímá, organizací, kde člověk pracuje apod.,
- samostatné vyhledání zdrojů – vyhledávače (např. Google), akademické zdroje, internetové aplikace (např. AWS Public Datasets) (Yau, 2011).

6.2 Ukázka vizualizace dat

Pro tuto praktickou část práce byla data získána z Českého statistického úřadu. Ve veřejných databázích byl použit vyhledávací dotaz „*knihovny*“. Ve výsledcích vyhledávání bylo zobrazeno několik výsledků bez popisných údajů. Některé z nich obsahovaly pouze tabulky, jiné i sloupcové grafy. Pro účely vizualizace byla k dispozici data ke stažení ve formátu XLS (Microsoft Excel) a XML (Extensible Markup Language)²⁹. Byla stažena tabulka s daty obsahující tyto informace:

- počet knihoven,
- knihovní jednotky:
 - naučná literatura,
 - krásná literatura,
- elektronické dokumenty,
- počet registrovaných:
 - z toho do 15 let,
- návštěvníci celkem:
 - fyzické návštěvy,
 - návštěvníci on-line služeb,
- počet výpůjček za rok:
 - naučná literatura dospělým,
 - naučná literatura dětem,
 - krásná literatura dospělým
 - krásná literatura dětem,
 - elektronické dokumenty.

²⁹ značkovací jazyk

Knihovny																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																										
----------	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

Kód: KUL002D30201/3

Obrázek 13: Data stažená z ČSÚ

Pro období roku 2015 je v dostupné tabulce uveden i srovnávací rok 2011. V tabulce není uveden počet registrovaných nad 15 let, číslo můžeme získat odečtením uživatelů do 15 let od celkového počtu registrovaných osob. Stažený soubor obsahuje i metainformace a statistické ukazatele. Sloupce tabulky označují jednotlivé kraje České republiky. Zkoumaným rokem je pro naše účely rok 2015, rok 2011 uvedený v tabulce k není potřeba, tudíž tabulku upravíme tak, aby informace v ní obsažené se vztahovaly pouze k údajům z roku 2015.

Jelikož se jedná pouze o jednoduchou tabulku, musíme věnovat důkladnou pozornost jednotlivým číslům. Existují určité vizuální prvky, které mohou jednoduše upozornit čtenáře na určité číslo ještě předtím, než čtenář věnuje tabulce vědomou pozornost. Tidwell (2011) takové prvky označuje jako „*preattentive variables*”³⁰. Jak již bylo uvedeno, vizualizace by měla fungovat efektivně, aby pomohla uživatelům odhalovat data. V knize proto poskytuje Tidwell několik příkladů těchto „*preattentive variables*“, jeden z nich je aplikován v následující tabulce.

³⁰ proměnné mající naši předběžnou pozornost

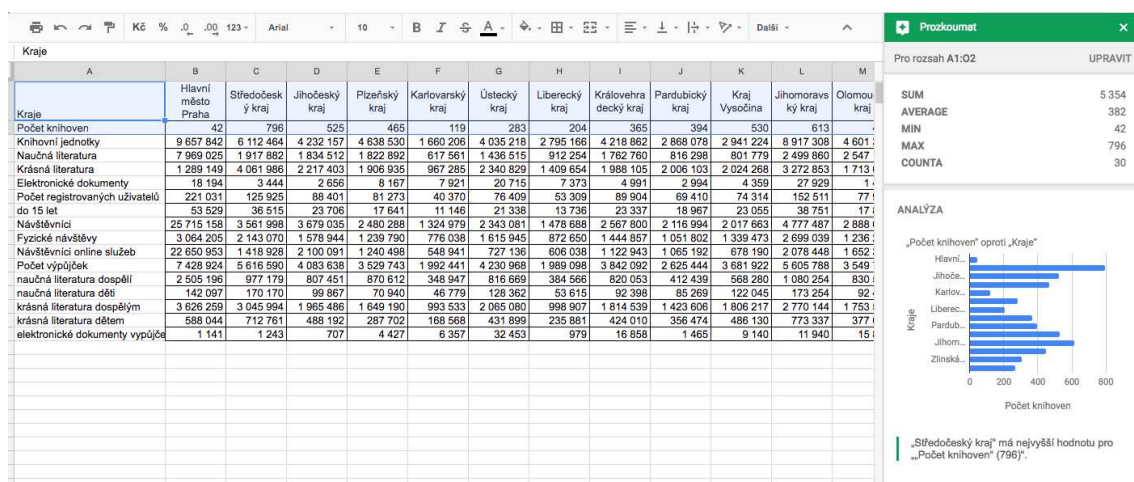
		Česká republika	Hlavní město Praha	Středočeský kraj	Jihočeský kraj	Plzeňský kraj	Karlovarský kraj	Ústecký kraj	Liberecký kraj	Královéhradecký kraj	Pardubický kraj	Kraj Vysočina	Jihomoravský kraj	Olomoucký kraj	Zlínský kraj	Moravskoslezský kraj
Počet knihoven	2015	5 354	42	796	525	465	119	283	204	365	394	530	613	448	303	287
Knihovny	2015	84 831 670	9 657 842	6 112 464	4 232 157	4 638 530	1 660 206	4 035 218	2 795 166	4 218 862	2 868 078	2 941 224	8 917 308	4 601 348	3 006 846	5 146 421
z toho	2015	27 930 557	7 969 025	1 917 882	1 834 512	1 822 892	617 561	1 436 515	912 254	1 762 760	816 298	801 779	2 499 860	2 547 173	967 136	2 024 910
naučná	2015	29 737 403	1 289 149	4 061 986	2 217 403	1 906 935	967 285	2 340 829	1 409 654	1 988 105	2 006 103	2 024 268	3 272 853	1 713 013	1 934 620	2 605 200
krásná literatura	2015	121 928	16 194	3 444	2 656	8 167	7 921	20 715	7 373	4 991	2 994	4 359	27 929	1 432	3 538	9 215
elektronické	2015	1 412 487	221 031	125 925	88 401	81 273	40 370	76 409	53 309	89 904	69 410	74 314	152 511	77 997	90 319	171 314
Počet registrovaných uživatelů	2015	373 182	53 529	36 515	23 706	17 641	11 146	21 338	13 736	23 337	18 967	23 055	38 751	17 861	25 696	47 904
z toho	2015	62 912 947	25 715 158	3 561 998	3 679 035	2 480 288	1 324 979	2 343 081	1 478 688	2 567 800	2 116 994	2 017 663	4 777 487	2 888 605	3 149 626	4 811 345
Návštěvnost	2015	23 623 299	3 064 205	2 143 070	1 578 944	1 239 790	776 038	1 615 945	872 650	1 444 857	1 051 802	1 339 473	2 699 039	1 236 233	1 651 738	2 509 515
do 15 let	2015	39 289 648	22 650 953	1 418 928	2 100 081	1 245 488	548 941	727 136	506 038	1 122 945	1 065 192	678 190	2 078 448	1 052 372	1 496 088	1 901 830
navštěvnost	2015	60 044 935	7 428 924	5 616 590	4 083 638	3 529 743	1 992 441	4 230 968	1 989 098	3 842 092	2 625 444	3 681 922	5 605 788	3 548 726	4 635 412	7 263 149
Počet výpůjček	2015	12 600 691	2 505 196	977 179	807 451	870 612	348 947	816 669	384 566	820 053	412 439	568 280	1 080 254	830 586	837 321	1 341 138
naučná literatura	2015	1 600 520	142 097	170 170	99 867	70 940	46 779	128 362	53 615	92 398	85 269	122 045	173 254	92 435	110 096	213 193
z toho	2015	29 847 438	3 626 259	3 045 994	1 965 486	1 649 190	993 533	2 065 080	998 907	1 814 539	1 423 606	1 806 217	2 770 144	1 753 565	2 350 089	3 584 849
krásná literatura	2015	6 680 341	588 044	712 761	488 192	287 702	168 568	431 899	235 881	424 010	356 474	486 130	773 337	377 087	469 534	880 742
elektronické	2015	122 910	1 141	1 243	707	4 427	6 357	32 453	979	16 858	1 465	9 140	11 940	15 832	2 598	17 819

Obrázek 14: Aplikace preattentive variables

Jestliže uživatel hledá určitou hodnotu (např. nejvyšší, nejnižší číslo), změněním některé z vizuálních funkcí tabulky se výrazně zkrátí čas k nalezení požadované hodnoty. Na uvedeném příkladu vidíme, že podbarvení určitých buněk tabulky upoutá okamžitě čtenářovu pozornost. Zjistíme, že zatímco ve Středočeském kraji se nachází největší počet knihoven, nejnavštěvovanějšími jsou knihovny v Praze, přestože se zde nachází naopak nejméně poboček. Také počet výpůjček je v Praze nejvyšší v celé České republice.

6.3 Google Spreadsheets

Podobně jako Microsoft Excel, nabízí základní typy grafů i cloudový Google Spreadsheets. Základní tvorba tabulky je snadná. V tabulce jednoduše označíme rozsah dat, která nás zajímají a pomocí panelu Explore (Prozkoumat) v pravém dolním rohu můžeme vizualizovat data. Automaticky se také objeví sumarizace údajů – celkový počet, průměrný počet, minimální a maximální hodnota.



Obrázek 15: Práce s daty v Google Spreadsheets

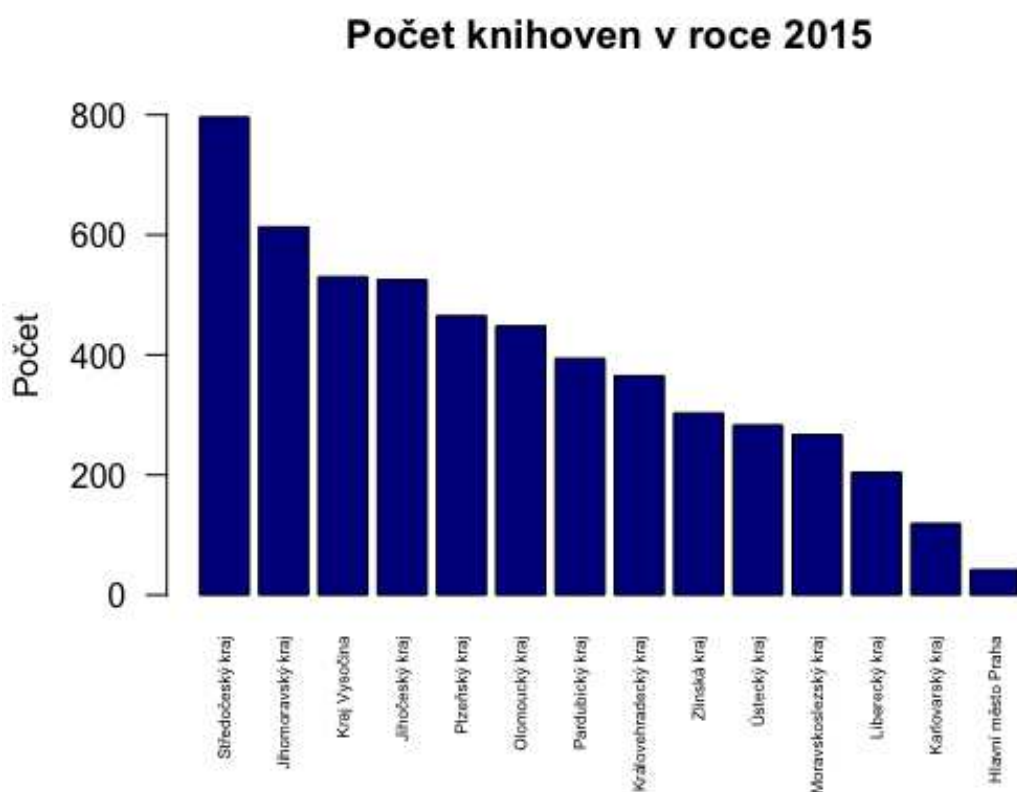
6.4 Vizualizace v R

R je programovací jazyk s otevřeným zdrojovým kódem (open-source) používaný pro statistickou analýzu dat a jejich následné grafické zobrazení. Vizualizace dat byla

provedena v *RStudio* – integrovaném vývojovém prostředí pro jazyk R. Importovat data lze například z Excelu nebo CSV. Byl proveden import jednoduché tabulky s názvy jednotlivých krajů a počtem knihoven v nich. Importovaná data byla zobrazena ve sloupcovém grafu, ve kterém se na ose X nachází názvy krajů, na ose Y počet knihoven. Graf byl seřazen od nejvyššího počtu knihoven do nejnižšího. V konzoli byl spuštěn následující kód:

```
Pocetknihoven2 <- Pocetknihoven2[order(Pocetknihoven2$Počet,decreasing = TRUE),]
barplot(Pocetknihoven2$Počet, col=c("darkblue"), ylab = "Počet", main = "Počet
knihoven v roce 2015", names.arg=Pocetknihoven2$Kraj, las = 2, cex.names = 0.5,
ylim=c(0, max = 800))
```

Z kódu je patrné, že k popiskům sloupců jsou přiřazeny jednotlivé kraje, písmo muselo být zmenšeno kvůli dlouhým názvům krajů (*cex.names*), maximální rozsah osy Y byl stanoven na 800 (*ylim=c(0, max = 800)*).



Obrázek 16: Vizualizace počtu knihoven v *RStudio*

6.5 Vizualizace v Pythonu

Python je programovací jazyk vyvíjen jako open-source projekt. Pro práci s daty je vhodná instalace například těchto knihoven: IPython Notebook, Matplotlib, Numpy, Pandas, Pillow, Pygments. *IPython Notebook* je nástroj běžící online v prohlížeči, který usnadňuje interaktivní práci v Pythonu a umožňuje zobrazení vizualizací. Vhodným formátem pro načtení dat pomocí Pandas je CSV.

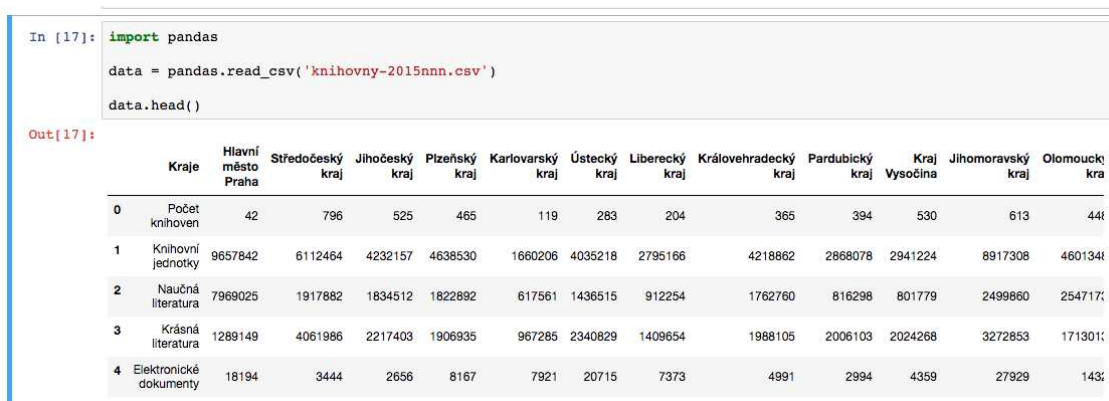
Data načteme pomocí kódu:

```
import pandas

data = pandas.read_csv('názevsouboru.csv')

data.head()
```

Metoda `head()` se používá pro zobrazení prvních řádků tabulky, abychom viděli, zda jsou data správně naformátovaná a výpis dat nebyl zbytečně dlouhý.



In [17]:

```
import pandas

data = pandas.read_csv('knihovny-2015nnn.csv')

data.head()
```

Out[17]:

	Kraje	Hlavní město Praha	Středočeský kraj	Jihočeský kraj	Plzeňský kraj	Karlovarský kraj	Ústecký kraj	Liberecký kraj	Královéhradecký kraj	Pardubický kraj	Kraj Vysočina	Jihomoravský kraj	Olomoucký kraj
0	Počet knihoven	42	796	525	465	119	283	204	365	394	530	613	441
1	Knihovní jednotky	9657842	6112464	4232157	4638530	1660206	4035218	2795166	4218862	2868078	2941224	8917308	4601341
2	Naučná literatura	7969025	1917882	1834512	1822892	617561	1436515	912254	1762760	816298	801779	2499860	2547171
3	Krásná literatura	1289149	4061986	2217403	1906935	967285	2340829	1409654	1988105	2006103	2024268	3272853	1713011
4	Elektronické dokumenty	18194	3444	2656	8167	7921	20715	7373	4991	2994	4359	27929	1432

Obrázek 17: Načtení dat pomocí Pandas v IPython Notebook

Pro zobrazení jen určitých údajů, které nás zajímají zadáme následující kód:

```
data['název sloupce']
```

```
data['Elektronické dokumenty']

Kraje
0 Hlavní město Praha 18194
1 Středočeský kraj 3444
2 Jihočeský kraj 2656
3 Plzeňský kraj 8167
4 Karlovarský kraj 7921
5 Ústecký kraj 20715
6 Liberecký kraj 7373
7 Královéhradecký kraj 4991
8 Pardubický kraj 2994
9 Kraj Vysočina 4359
10 Jihomoravský kraj 27929
11 Olomoucký kraj 1432
12 Zlínský kraj 3538
13 Moravskoslezský kraj 8215
Name: Elektronické dokumenty, dtype: int64
```

Obrázek 18: Ukázka načtení dat ze sloupce „Počet elektronických dokumentů“

V tomto prostředí lze pracovat i s aritmetickými operacemi, jednoduše zjistíme například poměr počtu registrovaných uživatelů na počet knihoven v jednotlivých krajích.

```
data['Počet registrovaných uživatelů'] / data['Počet knihoven']
```

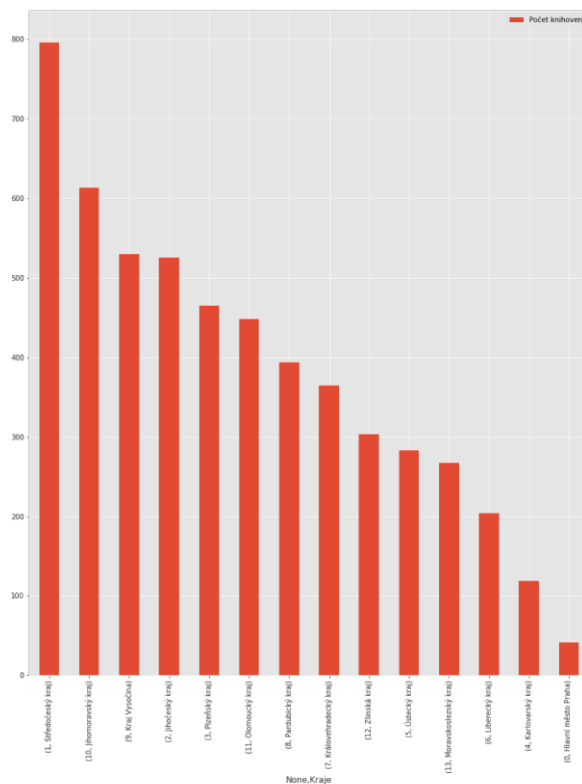
Kraje	
0	Hlavní město Praha 5262.642857
1	Středočeský kraj 158.197236
2	Jihočeský kraj 168.382857
3	Plzeňský kraj 174.780645
4	Karlovarský kraj 339.243697
5	Ústecký kraj 269.996466
6	Liberecký kraj 261.318627
7	Královéhradecký kraj 246.312329
8	Pardubický kraj 176.167513
9	Kraj Vysočina 140.215094
10	Jihomoravský kraj 248.794454
11	Olomoucký kraj 174.100446
12	Zlínská kraj 298.082508
13	Moravskoslezský kraj 641.625468

Obrázek 19: Ukázka aritmetická operace v IPython Notebook

Vizualizace počtu knihoven byla vytvořena následujícím kódem:

```
data[['Počet knihoven']].sort_values(by='Počet knihoven', ascending=False).plot(kind='bar')
```

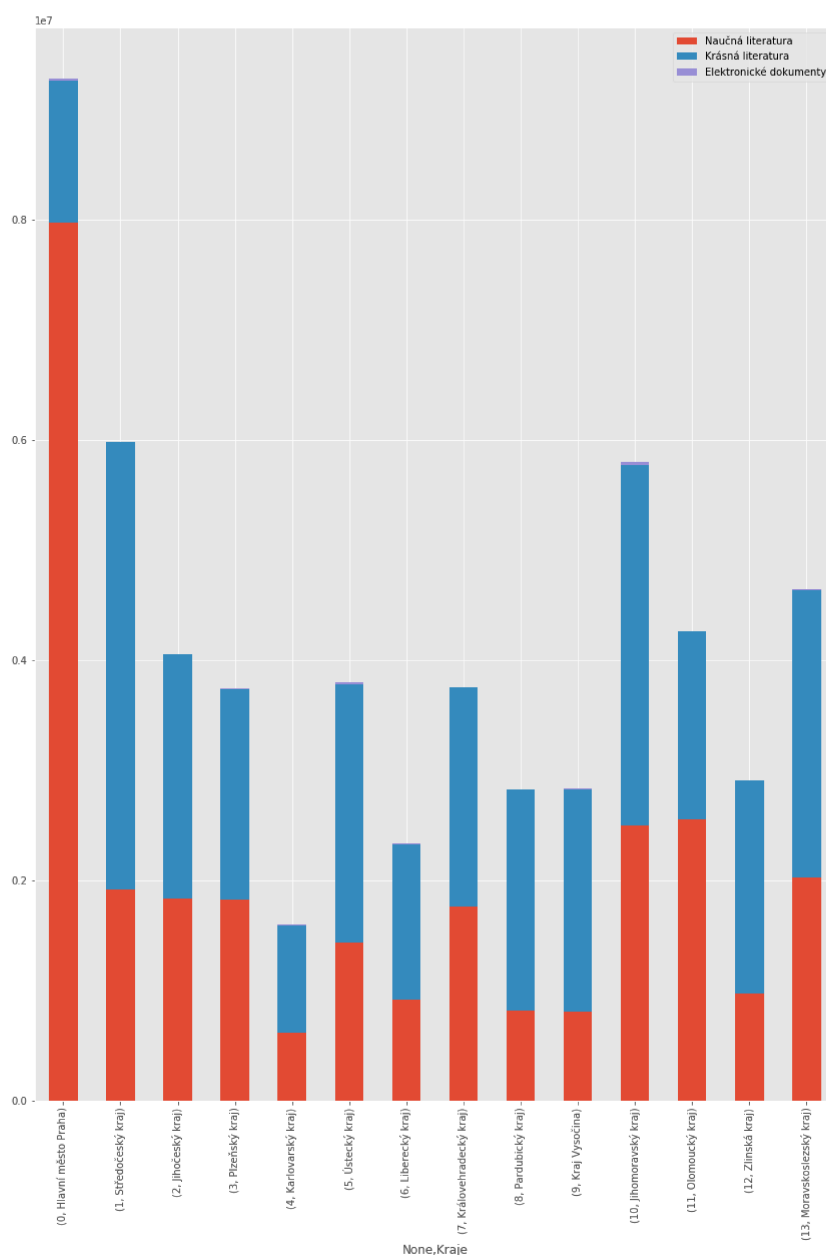
Opět byla data seřazena od nejvyššího počtu po nejnižší (sort_values), byl zvolen sloupčový graf.



Obrázek 20: Vizualizace počtu knihoven v IPython Notebook

Další ukázkou vizualizace je skládaný sloupcový graf se zobrazením poměru jednotlivých typů dokumentů, které jsou k dispozici v knihovnách. Vytvoříme jej pomocí kódu:

```
data[['Naučná literatura', 'Krásná literatura', 'Elektronické dokumenty']].plot(kind='bar', stacked=True)
```



Obrázek 21: Podíl jednotlivých typů dokumentů v knihovnách v IPython Notebook

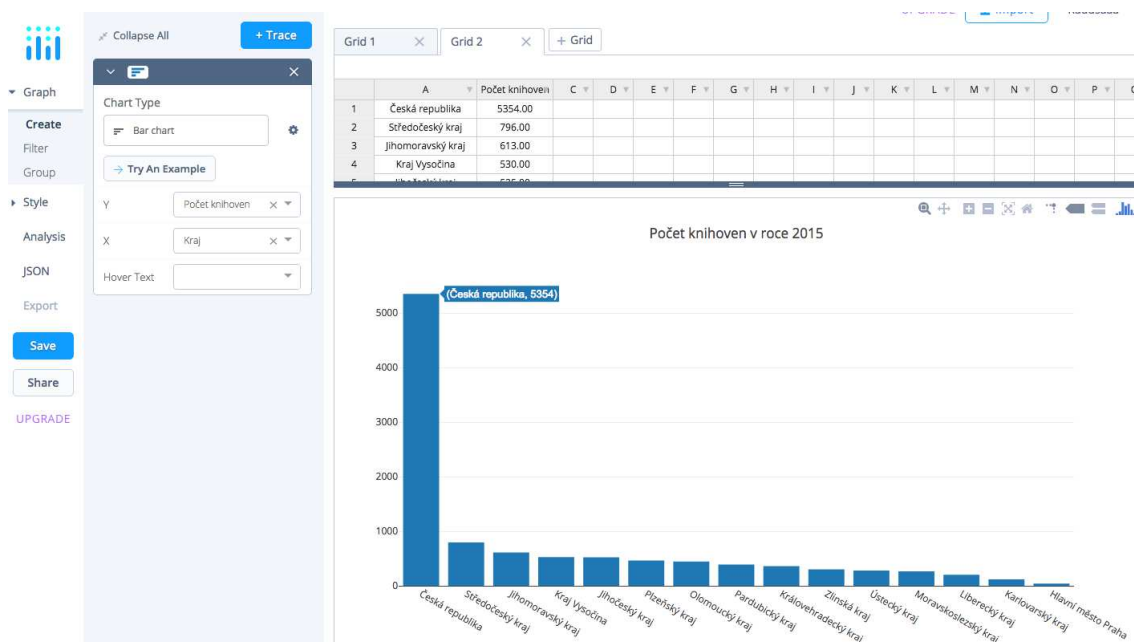
6.6 Vizualizace pomocí Out-of-the-box software

Následující vizualizace jsou vytvořeny bez nutnosti znalosti programování pomocí nástrojů dostupných online.

6.6.1 Plotly

Plotly³¹ je online nástroj, pomocí kterého lze vytvářet grafy, prezentace či nástěnky. Poskytuje také analytické a statistické nástroje a knihovny pro Python, R, Matlab a další jazyky. Je vhodná jak pro jednotlivce, tak i pro spolupráci v týmu a firmy. Plotly můžeme také použít pro tvorbu vizuálně atraktivnějších interaktivních grafů v IPythonu. Data můžeme importovat v Excelu, CSV nebo pomocí SQL. Použití tohoto nástroje je zdarma za předpokladu veřejného publikování výsledných vizualizací. V případě potřeby privátních grafů (např. z důvodu zobrazení citlivých dat) je služba dostupná za poplatek.

V tomto případě byla data naimportována ve formátu XLS. Práce v tomto online nástroji je intuitivní. V horní oblasti vidíme nahraná data, vlevo si můžeme zvolit typ grafu, co bude zobrazeno na ose X a Y a další prvky, kterými můžeme výsledný graf měnit (barvy, text). Po najetí kurzorem myši na daný sloupec grafu se zobrazí přesná číselná hodnota.



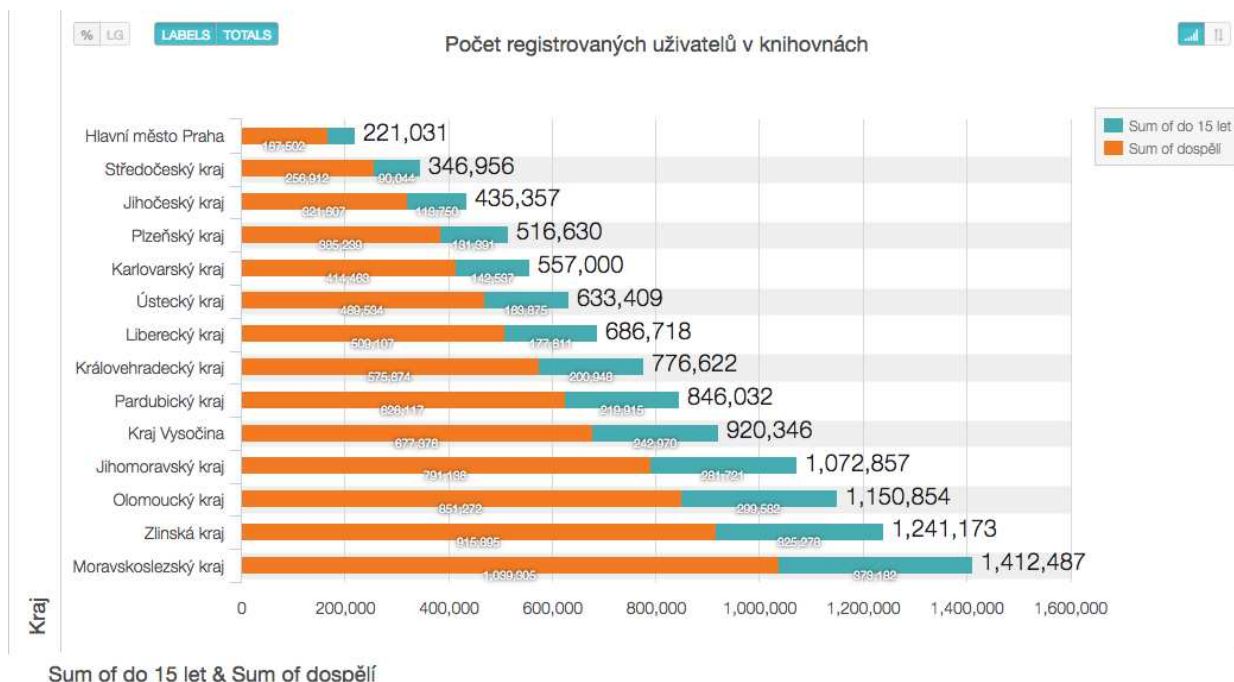
Obrázek 22: Tvorba vizualizace dat v Plotly

³¹ dostupné z: <https://plot.ly/>

6.6.2 DataHero

DataHero³² je další z online nástrojů pro vizualizaci dat. Umožňuje shromažďovat data z cloudových služeb a vytvářet grafy. Pro tvorbu vizualizací nejsou potřeba žádné technické znalosti, proto je vhodný pro použití jak pro jednotlivce, tak i v celých týmech bez nutnosti IT odborníků. DataHero má předem připravené propojení ke cloudovým službám (Google Drive, Dropbox apod.) a zajistí tak okamžitý přístup k datům a následnou práci s nimi. Pracuje se v intuitivním rozhraní na bázi „*drag-and-drop*“ (táhni a pusť). Základní verze je zdarma, pro více možností je nutné měsíční předplatné.

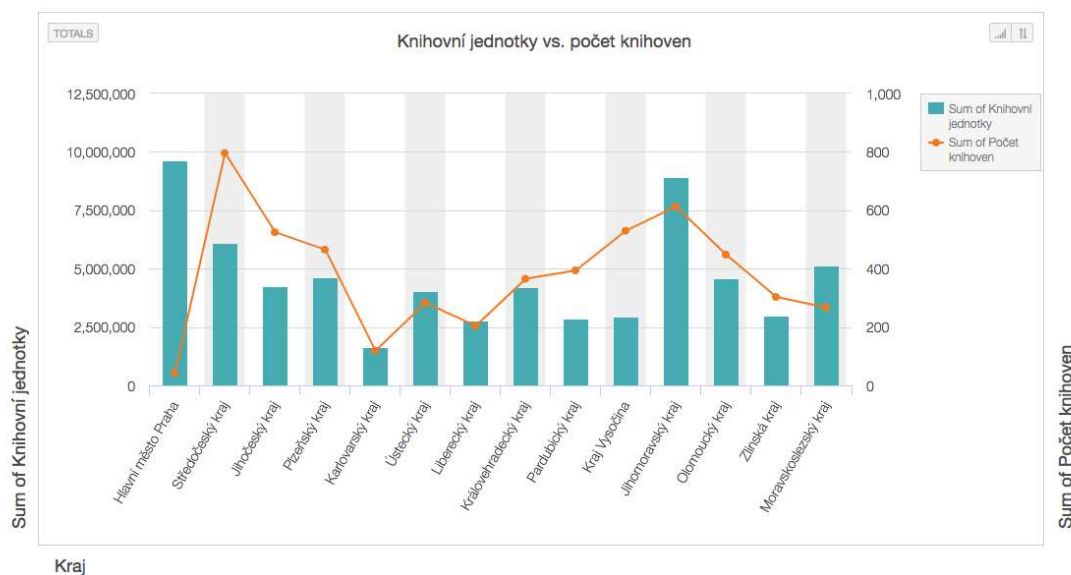
Pro práci s tímto nástrojem byla data nahrána ve formátu XLS. Na základě nahraných dat DataHero automaticky navrhuje, jaké typy grafů jsou vhodné k zobrazení. Lze měnit jak typ grafu, tak i celkovou vizuální podobu (barva, písmo, velikost). Byla provedena vizualizace počtu registrovaných uživatelů v knihovnách pomocí skládaného pruhového grafu, kdy vidíme rozdělení na uživatele do 15 let a dospělé. Při najetí kurzoru myši na jednotlivý pruh se zobrazí souhrnné informace – název kraje, počet uživatelů a vyjádření v procentech.



Obrázek 23: Vizualizace počtu uživatelů knihoven v DataHero

³² dostupné z: <https://datahero.com/>

Dále byla provedena vizualizace počtu knihovních jednotek a počtu knihoven v daných krajích pomocí složeného grafu – sloupcového a spojnicového.



Obrázek 24: Složený graf v DataHero

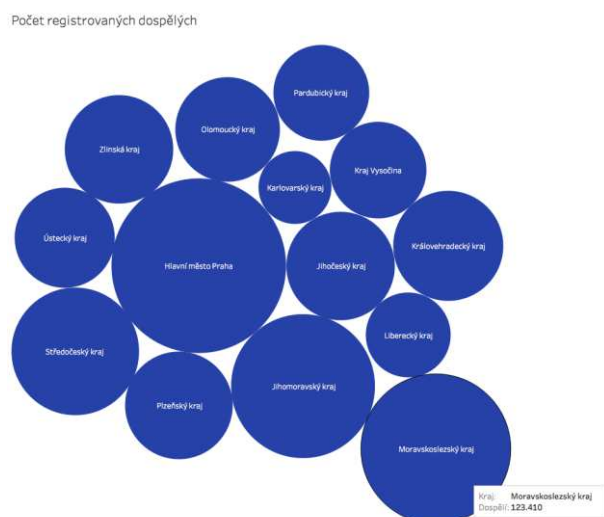
Sloupce označují knihovní jednotky a spojnice počet knihoven dle krajů. Z vizualizace vyplývá, že přestože má Praha nejmenší počet knihoven je zde dostupných nejvíce knihovních jednotek v celé České republice.

6.6.3 Tableau Software

Tato společnost poskytuje nástroje pro vytváření interaktivních datových vizualizací se zaměřením na business intelligence. Pro tvorbu vizualizace bylo použito Tableau Public³³, které je zdarma, ostatní verze jsou placené. Verze zdarma funguje na podobném principu jako Plotly, kdy vytvořené vizualizace se stávají veřejnými. Na rozdíl od předchozích nástrojů je u tohoto softwaru nutné stažení aplikace do počítače. Rozhraní je opět intuitivní a na bázi „drag-and-drop“.

Po stažení aplikace byla data nahrána ve formátu XLS. Proběhla tvorba bublinového grafu počtu registrovaných dospělých a dětí do 15 let v jednotlivých krajích České republiky. Po najetí kurzoru myši na jednotlivou bublinu se zobrazí přesný počet uživatelů knihoven představujících danou bublinu.

³³ dostupné z: <https://public.tableau.com/>



Obrázek 25: Vizualizace počtu registrovaných dospělých v Tableau Public



Obrázek 26: Vizualizace počtu registrovaných dětí do 15 let v Tableau Public

6.7 Shrnutí praktické části

V praktické části byla demonstrována tvorba vizualizace v různých nástrojích. Byla provedena úprava veřejně dostupných dat v tabulkovém formátu tak, aby výsledná prezentace dat byla přehledná, čitelná a jasnější. Byly představeny také nástroje, u nichž není potřeba žádná předchozí znalost programování nebo technických základů. Vizualizace v nich představuje moderní alternativu k často používanému MS Excel. Práce s těmito nástroji je intuitivní, software je uživatelsky přívětivý a získáme k němu přístup odkudkoli a z jakéhokoli zařízení, protože většina z těchto moderních aplikací je dostupná online v prohlížeči. Výsledné vizualizace lze tedy i jednoduše sdílet a

publikovat online. Z těchto důvodů jsou tyto nástroje užitečné jak pro jednotlivce, tak i pro týmovou spolupráci. Jsou vhodné pro osobní i firemní použití, kdy za poměrně krátkou dobu je člověk schopen efektivně vizualizovat a prezentovat data. V posledních letech se nástroje k vizualizaci rozvíjí a vznikají stále nové zejména v důsledku poptávky na straně zákazníků a vyhovění jejich požadavkům. Praktická část také ukázala, že MS Excel rozhodně nemusí být jedinou volbou nástroje k tvorbě grafu, jestliže nedisponujeme ani základy programování. V představených uživatelsky přívětivějších nástrojích je často tvorba rychlejší a výsledné vizualizace jsou esteticky působivější.

7 Závěr

V teoretické části bakalářské práce *Vizualizace jako metoda prezentace dat* byl nastíněn historický vývoj vizualizace dat s důrazem na jednotlivé milníky a významné osobnosti. Zvláštní pozornost byla věnována i současnému vývoji trendů v oblasti prezentace a komunikace dat publiku. Byla rozebrána problematika terminologie jednotlivých pojmů a provedeno srovnání definic vizualizace dat s poukázáním na rozdílnost výkladu jednotlivých autorů a stanovením takových kritérií, aby mohlo být dílo považováno za vizualizaci dat. Byl popsán samotný proces vizualizace a její fungování a představeny důvody, proč je tato metoda zobrazování dat efektivní a účinná. Byly uvedeny způsoby, jakými lze vytvářet vizualizace různých typů dat za pomoci užitečných softwarových nástrojů a v praktické části předvedena tvorba vizualizace za použití těchto nástrojů. Cíl práce – potvrzení, že vizualizace dat je efektivní způsob, jakým můžeme prezentovat data, která jsou v surovém stavu nepřehledná – byl splněn.

Na základě dostupné literatury bylo zjištěno, že oblast vizualizace dat nelze lehce ohraničit a definovat. Střetávají se v ní odborníci z různých oblastí a každý z nich postupuje při tvorbě vizualizace jinak a používá ji k jiným účelům. I z tohoto důvodu nelze pokládat téma této práce za vyčerpané. V práci byl kladen důraz zejména na uvedení do problematiky vizualizace s poukázáním na aktuální vývoj a představení nástrojů vhodných pro vizualizaci. V celé práci se prolíná téma efektivity a hledání souvislostí v prezentovaných datech.

Jak poznamenává Nathan Yau (2011), nesmíme zapomínat, že data jen reprezentují skutečný život a svět kolem nás. Jestliže se s nimi naučíme správně pracovat, můžeme pomocí nich vyprávět příběhy, o kterých lidé ani netuší, že existují. Lidstvo v dnešní době produkuje takové množství dat, jako nikdy předtím a v reakci na ně vznikají stále nové nástroje pro práci s nimi. Z těchto důvodů existují vhodné podmínky pro to, aby tato data byla zpracována a pomocí jejich vizualizace dána do kontextu, jak na individuální úrovni, tak i v mnohem větším měřítku pokrývajícím celý svět.

Mnohé příklady z minulosti již potvrdily, že vizualizace jako metoda prezentace dat je efektivní, proto bude nesmírně zajímavé sledovat její budoucí vývoj, ve kterém máme téměř neomezené možnosti. Jak poznamenává John W. Tukey: „*Největší hodnota obrazu spočívá v jeho síle upozornit nás vidět to, co nikdy neočekáváme.*“

Seznam použité literatury

1. BAILEY, Jefferson a Lily PREGILL. Speak to the Eyes: The History and Practice of Information Visualization. *Art Documentation: Journal of the Art Libraries Society of North America* [online]. 2014, **33**(2), 168-191 [cit. 2017-07-15]. DOI: 10.1086/678525. ISSN 0730-7187. Dostupné z: <http://www.journals.uchicago.edu/doi/10.1086/678525>
2. BATEMAN, Scott, Regan L. MANDRYK, Carl GUTWIN, Aaron GENEST, David MCDINE a Christopher BROOKS. Useful junk?: the effects of visual embellishment on comprehension and memorability of charts. In: *Proceedings of the 28th international conference on Human factors in computing systems - CHI '10* [online]. New York, USA: ACM Press, 2010, s. 2573-2582 [cit. 2017-07-15]. DOI: 10.1145/1753326.1753716. ISBN 978-1-60558-929-9. Dostupné z: <http://portal.acm.org/citation.cfm?doid=1753326.1753716>
3. BENEŠ, Bedřich, Petr FELKEL, Jiří SOCHOR a Jiří ŽÁRA. *Vizualizace*. Praha: České vysoké učení technické, 1997. ISBN 80-010-1582-3.
4. BOLLEN, Johan, Herbert VAN DE SOMPEL, Aric HAGBERG, Luis BETTENCOURT, Ryan CHUTE, Marko A. RODRIGUEZ a Lyudmila BALAKIREVA. Clickstream Data Yields High-Resolution Maps of Science. *PLoS ONE* [online]. 2009, **4**(3), 1-11 [cit. 2017-07-15]. DOI: <https://doi.org/10.1371/journal.pone.0004803>. Dostupné z: <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0004803>
5. CARD, Stuart K., Jock D. MACKINLAY a Ben SHNEIDERMAN. *Readings in information visualization: using vision to think*. San Francisco, California: Morgan Kaufmann Publishers, c1999. ISBN 978-1-5586053-3-6.
6. FRIENDLY, Michael. Visions and Re-Visions of Charles Joseph Minard. *Journal of Educational and Behavioral Statistics* [online]. 2002, **27**(1), 31-51 [cit. 2017-06-29]. ISSN 10769986. Dostupné z: <http://www.jstor.org/stable/3648145>

7. FRIENDLY, Michael. A Brief History of Data Visualization. CHEN, Chunhou, Wolfgang HÄRDLE a Antony UNWIN, ed. *Handbook of Data Visualization* [online]. Berlin, Heidelberg: Springer-Verlag Berlin Heidelberg, c2008, s. 15-56 [cit. 2017-06-29]. DOI: 10.1007/978-3-540-33037-0_2. ISBN 978-3-540-33037-0. Dostupné z: http://link.springer.com/10.1007/978-3-540-33037-0_2
8. FRY, Ben. *Visualizing data*. Sebastopol: O'Reilly Media, c2008. ISBN 978-0-596-51455-6.
9. GERSHON, Nahum a Ward PAGE. What storytelling can do for information visualization. *Communications of the ACM* [online]. 2001, **44**(8), 31-37 [cit. 2017-07-10]. DOI: 10.1145/381641.381653. ISSN 0001-0782. Dostupné z: <http://portal.acm.org/citation.cfm?doid=381641.381653>
10. HistoryWired: A Few of Our Favorite Things. *National Museum of American History* [online]. Washington, D.C.: The Smithsonian Institution [cit. 2017-07-17]. Dostupné z: <http://americanhistory.si.edu/exhibitions/history-wired>
11. KERREN, Andreas, John T. STASKO, Jean-Daniel FEKETE a Chris NORTH, ed. *Information visualization: human-centered Issues and perspectives* [online]. Berlin: Springer-Verlag Berlin Heidelberg, 2008 [cit. 2017-06-29]. ISBN 978-354-0709-558. Dostupné z: <http://www.springer.com/gp/book/9783540709558>
12. KOSARA, Robert. Visualization Criticism - The Missing Link Between Information Visualization and Art. In: *2007 11th International Conference Information Visualization (IV '07)* [online]. Zurich: IEEE, 2007, s. 631-636 [cit. 2017-07-05]. DOI: 10.1109/IV.2007.130. ISBN 0-7695-2900-3. ISSN 1550-6037. Dostupné z: <http://ieeexplore.ieee.org/document/4272046/>
13. KOSARA, Robert a Jock MACKINLAY. Storytelling: The Next Step for Visualization. *Computer* [online]. 2013, **46**(5), 44-50 [cit. 2017-07-10]. DOI: 10.1109/MC.2013.36. ISSN 0018-9162. Dostupné z: <http://ieeexplore.ieee.org/document/6412677/>

14. KRUM, Randy. *Cool infographics: effective communication with data visualization and design*. Indianapolis: John Wiley & Sons, c2014. ISBN 978-1-118-58230-5.
15. LAUMANS, Joel. *An introduction to visualizing data* [online]. 2009 [cit. 2017-07-05]. Dostupné z: <http://www.babaksohrabi.com/Files/TextNews/AnIntroductionToVisualizingData.pdf>
16. LIMA, Manuel. *Visual complexity: mapping patterns of information*. New York: Princeton Architectural Press, c2011. ISBN 978-1-56898-936-5.
17. LIU, Ling a M. Tamer ÖZSU, ed. *Encyclopedia of database systems* [online]. New York: Springer, c2009 [cit. 2017-06-29]. ISBN 978-0-387-39940-9. Dostupné z: <http://www.springer.com/in/book/9780387355443>
18. LORENSEN, Bill. On the death of visualization: can it survive without customers? In: *NIH/NSF Proc. Fall 2004 Workshop Visualization Research Challenges* [online]. Los Alamitos: IEEE Press, 2004, s. 1-5 [cit. 2017-07-10]. Dostupné z: https://erie.nlm.nih.gov/evc/meetings/vrc2004/position_papers/lorensen.pdf
19. MA, I. LIAO, J. FRAZIER, Helwig HAUSER a Helen-Nicole KOSTIS. Scientific Storytelling Using Visualization. *IEEE Computer Graphics and Applications* [online]. 2012, **32**(1), 12-19 [cit. 2017-07-15]. DOI: 10.1109/MCG.2012.24. ISSN 0272-1716. Dostupné z: <http://ieeexplore.ieee.org/document/6111347/>
20. MANOVICH, Lev. Cultural Analytics: Visualizing Cultural Patterns in the Era of “More Media”. In: *Lev Manovich* [online]. 2009 [cit. 2017-07-29]. Dostupné z: <http://manovich.net/index.php/projects/cultural-analytics-visualizing-cultural-patterns>
21. MAREK, Tomáš. *Efektivní vizualizace dat se zaměřením na základní typy grafů* [online]. Brno, 2014 [cit. 2017-06-29]. Dostupné z: http://is.muni.cz/th/362075/ff_m/. Diplomová práce. Masarykova univerzita, Filozofická fakulta. Vedoucí práce Jan Boček.

22. MAZZA, Riccardo. *Introduction to information visualization* [online]. London: Springer-Verlag London Limited, c2009 [cit. 2017-07-02]. ISBN 978-1-84800-219-7. Dostupné z: <http://www.springer.com/la/book/9781848002180>
23. MCCORMICK, Bruce H., Thomas A. DEFANTI a Maxine D. BROWN, ed. Visualization in scientific computing. *ACM SIGBIO Newsletter* [online]. 1988, **10**(1), 15-21 [cit. 2017-06-29]. DOI: 10.1145/43965.43966. ISSN 0163-5697. Dostupné z: <http://portal.acm.org/citation.cfm?doid=43965.43966>
24. MEIRELLES, Isabel. *Design for information: an introduction to the histories, theories, and best practices behind effective information visualizations* [online]. Beverly: Rockport Publishers, 2013 [cit. 2017-07-29]. ISBN 978-159-2538-065. Dostupné z: <https://ebookcentral.proquest.com/>
25. RENDGEN, Sandra, Paolo CIUCCARELLI, Richard Saul WURMAN, Simon ROGERS a Nigel HOLMES, WIEDEMANN, Julius, ed. *Information graphics*. Italy: TASCHEN, c2012. ISBN 978-3-8365-2879-5.
26. SAXL, Ivan a Lucia ILUCOVÁ. Historie grafického zobrazování statistických dat. In: *ROBUST 2004*. Praha: Jednota českých matematiků a fyziků, 2004, s. 363-386. ISBN 80-7015-972-3.
27. SEGEL, Edward a Jeffrey HEER. Narrative Visualization: Telling Stories with Data. *IEEE Transactions on Visualization and Computer Graphics* [online]. 2010, **16**(6), 1139-1148 [cit. 2017-07-15]. DOI: 10.1109/TVCG.2010.179. ISSN 1077-2626. Dostupné z: <http://ieeexplore.ieee.org/document/5613452/>
28. SHAPLEY, Deborah. The da Vinci of Data. *The New York Times* [online]. 1998 [cit. 2017-07-20]. Dostupné z: <http://www.nytimes.com/1998/03/30/business/the-da-vinci-of-data.html>
29. SHIODE, Shino. Revisiting John Snow's map: network-based spatial demarcation of cholera area. *International Journal of Geographical Information Science* [online]. 2012, **26**(1), 133-150 [cit. 2017-06-29]. DOI: 10.1080/13658816.2011.577433. ISSN 1365-8816. Dostupné z: <http://www.tandfonline.com/doi/abs/10.1080/13658816.2011.577433>

30. SMICIKLAS, Mark. *The power of infographics: using pictures to communicate and connect with your audience*. Indianapolis: Que Pub., c2012. ISBN 9780789749499.
31. SPENCE, Robert. *Information visualization*. Third Edition. Switzerland: Springer International Publishing, 2014. ISBN 978-3-319-07340-8.
32. STEELE, Julie a Noah P. N. ILIINSKY, ed. *Beautiful visualization: looking at data through the eyes of experts*. Sebastopol: O'Reilly, c2010. ISBN 978-1-4493-7986-5.
33. TELEA, Alexandru C. *Data visualization: principles and practice*. Second edition. Boca Raton: Taylor & Francis Group, c2015. ISBN 978-1-4665-8526-3.
34. TIDWELL, Jenifer. *Designing interfaces*. 2nd ed. Sebastopol: O'Reilly, c2011. ISBN 978-144-9379-704.
35. TUFTE, Edward R. *The visual display of quantitative information*. 2nd ed. Cheshire, Connecticut: Graphics Press, c2001. ISBN 978-0-9613921-4-7.
36. TUKEY, John W. The Future of Data Analysis. *The Annals of Mathematical Statistics* [online]. 1962, **33**(1), 1-67 [cit. 2017-06-29]. DOI: 10.1214/aoms/1177704711. ISSN 00034851. Dostupné z: <http://projecteuclid.org/euclid.aoms/1177704711>
37. VAN WIJK, Jarke J. The Value of Visualization. In: *VIS 05. IEEE Visualization, 2005* [online]. IEEE, 2005, s. 79-86. DOI: 10.1109/VISUAL.2005.1532781. ISBN 0-7803-9462-3. Dostupné také z: <http://ieeexplore.ieee.org/document/1532781/>
38. WONG, Pak Chung a Jim THOMAS. Visual Analytics. *IEEE Computer Graphics and Applications* [online]. 2004, **24**(5), 20-21 [cit. 2017-07-07]. DOI: 10.1109/MCG.2004.39. ISSN 0272-1716. Dostupné z: <http://ieeexplore.ieee.org/document/1333623/>
39. YAU, Nathan. *Visualize this: the FlowingData guide to design, visualization, and statistics*. Indianapolis: Wiley Publishing, c2011. ISBN 978-0-470-94488-2.

Seznam obrázků

Obrázek 1: Rekonstrukce fresky nalezené v Catal Hüyük	10
Obrázek 2: Snowův plán Londýna v době cholery epidemie (1854)	12
Obrázek 3: Grafika Charlese Josepha Minarda, originál a anglický překlad	13
Obrázek 4: Přehled oblastí vizualizace	20
Obrázek 5: Konceptní pohled na proces vizualizace	23
Obrázek 6: Základní typy grafů	27
Obrázek 7: Pokročilé typy grafů	28
Obrázek 8: Styly vyprávění	30
Obrázek 9: Grafika od Nigela Holmese	35
Obrázek 10: čistý graf použitý v experimentu	35
Obrázek 11: Modrian vs. Rothko: Lev Manovich (2009)	36
Obrázek 12: Mapa vědních disciplín	38
Obrázek 13: Data stažená z ČSÚ	44
Obrázek 14: Aplikace preattentive variables	45
Obrázek 15: Práce s daty v Google Spreadsheets	45
Obrázek 16: Vizualizace počtu knihoven v RStudio	46
Obrázek 17: Načtení dat pomocí Pandas v IPython Notebook	47
Obrázek 18: Ukázka načtení dat ze sloupce „Počet elektronických dokumentů“	47
Obrázek 19: Ukázka aritmetická operace v IPython Notebook	48
Obrázek 20: Vizualizace počtu knihoven v IPython Notebook	48
Obrázek 21: Podíl jednotlivých typů dokumentů v knihovnách v IPython Notebook	49
Obrázek 22: Tvorba vizualizace dat v Plotly	50
Obrázek 23: Vizualizace počtu uživatelů knihoven v DataHero	51
Obrázek 24: Složený graf v DataHero	52
Obrázek 25: Vizualizace počtu registrovaných dospělých v Tableau Public	53
Obrázek 26: Vizualizace počtu registrovaných dětí do 15 let v Tableau Public	53